DOCUMENT RESUME

ED 181 344 CG 014 056

AUTHOR Wolman, Jean
TITLE Conducting Summative Evaluation (Cost-Impact Studies). Module 11.
INSTITUTION American Institutes for Research in the Behavioral Sciences, Palo Alto, Calif.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Jul 78
NOTE 104p.; For related documents see CG 014 049-057, ED 124 823 and ED 133 511. Reprint.
AVAILABLE FROM National Consortium Project, American Institutes for Research, P.O. Box 1113, Palo Alto, CA 94302 ($3.20)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Career Development; Counselors; *Evaluation Methods; Guidance Personnel; Inservice Programs; *Occupational Guidance; *Performance Criteria; Professional Development; Program Design; Program Effectiveness; *Program Evaluation; Secondary Education; *Summative Evaluation

ABSTRACT

This staff development module is part of one of three groups of career guidance modules developed, field-tested, and revised by a six-state consortium coordinated by the American Institutes for Research. This module is the eleventh in a series on developing a comprehensive career guidance program at the high school level, designed to aid guidance personnel responsible for developing student focused programs. The goal of this module is to provide users with an understanding of the importance of summative evaluation and the basic skills necessary for program evaluation including evaluation design, sample, instruments, administration, criteria, and decisions based on data results and analyses. The module format consists of an overview, goals, objectives, outline, time schedule, glossary, readings, skill development activities, and bibliography. A Coordinator's Guide is also included with detailed instructions for presenting the module in a workshop setting as well as the facilitator's roles and functions, and the criteria used in assessing the participants' achievement of module objectives. (Author/HLM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED181344

HARD COPY NOT AVAILABLE

MODULE 11

CONDUCTING SUMMATIVE EVALUATION

(Cost-Impact Studies)

Jean Wolman

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

A. I. R.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CG 014056

Developed at the American Institutes for Research, under support by the United States Office of Education, Department of Health, Education, and Welfare under Part C of the Vocational Education Act of 1963.

August 1975

Reprinted with minor revisions, July 1978

ISBN 0-89785-607-4

Illustrations in this module are by Jurgen Wolff. Editorial work is by Merlynn Bergen and Phyllis DuBois.

This Module builds on cooperative staff development activities of:

Mesa Public Schools; Mesa, Arizona,
Fremont Union High School District; Sunnyvale, California,
Santa Clara County Superintendent of Schools Office;
San Jose, California, and
American Institutes for Research, Palo Alto, California.

Support for these efforts was received through the United States Office of Education, Department of Health, Education, and Welfare under Part C of the Vocational Education Act of 1963 and Title III of the Elementary and Secondary Education Act of 1965. Points of view or opinions stated do not necessarily represent U.S.O.E. position or policy.

TABLE OF CONTENTS

Page

INTRODUCTION

Module Goal and Outcomes ---------- 1

Module Outline ---------- 2

Model and Module 11 ---------- 3

Glossary ---------- 5

TEXT

Designing a Summative Evaluation ---------- 7

Discussion - understanding the purpose and value of conducting summative evaluation ---- 17

Activity 1 - designing a summative evaluation -- 18

Selecting/Developing the Instruments ---------- 19

Activity 2 - developing evaluation instruments ---------- 25

Administering the Instruments ---------- 30

Discussion - administering the instruments ----- 33

Processing and Analyzing the Data ---------- 34

Activity 3 - preparing bar and line graphs ----- 42

Activity 4 - processing and analyzing the data ---------- 45

Relating Program Costs to Program Effects ---------- 47

Activity 5 - determining cost effectiveness ---- 51

Making Decisions Based on Results ---------- 54

Activity 6 - making decisions based on results ---------- 59

Summary ---------- 60

POSTASSESSMENT ---------- 61

APPLICATION ---------- 62

APPENDIX

Selected List of Test Publishers ---------- 67

References ---------- 69

COORDINATOR'S GUIDE

INTRODUCTION

MODULE GOAL AND OUTCOMES

Module Goal

To provide you with an understanding of the importance of summative evaluation and the basic skills needed to design and carry out an evaluation in your setting.

*Module Outcomes**

When you have successfully completed this module, you will be able to

1. State the purposes and value of conducting a summative evaluation.
2. Select an appropriate evaluation design and sample, given an evaluation situation.
3. Select, adapt, or adopt appropriate measurement instruments given particular evaluation needs.
4. List the major considerations involved in administering evaluation instruments.
5. Propose appropriate methods for processing, analyzing, and presenting evaluation data.
6. Identify important considerations in determining cost effectiveness.
7. State appropriate decisions based on evaluation results, given an evaluation situation.
8. Map out a plan for conducting a summative evaluation of your program.

*The criteria for attainment of the outcomes are available in the Coordinator's Guide for Module 11.

MODULE OUTLINE

Approximate Time	*Activity*	*Outcomes*
1 hour	Introduction Optional Introductory Activity. The Coordinator will explain the structure and purposes of the module. Questions and discussion included. Tape-slide presentation.	1
6 hours	Text Presentation of the important information related to conducting summative evaluations. Discussion sessions and practice activities interspersed.	1-7
20 minutes	Postassessment Assessment of your acquired knowledge and skills.	1-7
1 hour	Application Planning for the use of these abilities in your own setting.	8

A theme running throughout the comprehensive approach to developing career guidance programs is that of evaluation. Evaluation plays varied roles in the model (presented on the next page) in all four phases. The model in fact encourages repeated cycles of "formative" evaluation and decision making.

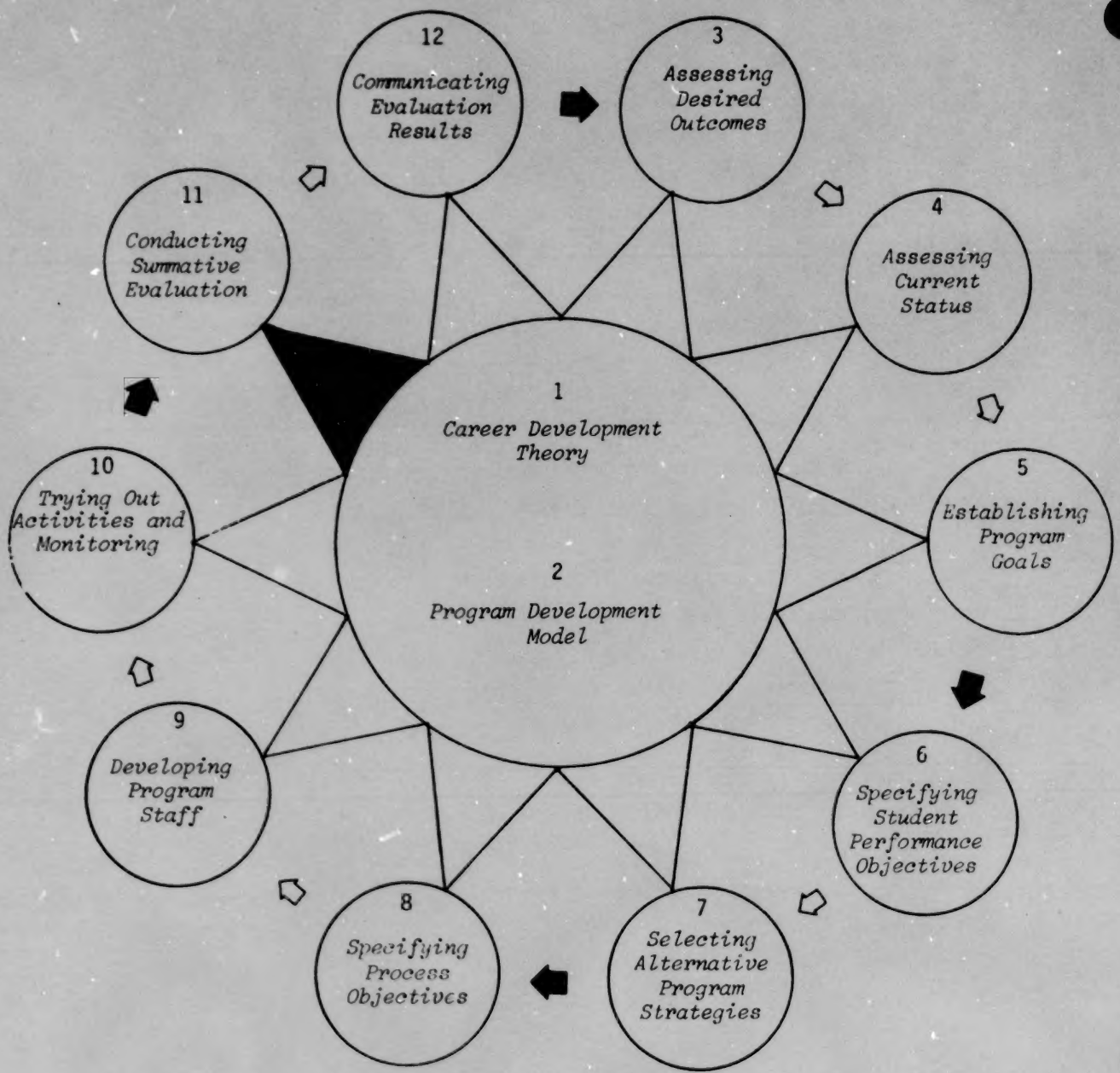
In the model, decisions are made at each phase of program development, and the evaluation results and decisions of all phases are interrelated. The successive phases are designed to answer the following questions:

1. What should the program goals and objectives be?
   (Phase 1 - Modules 3-5)
2. What strategies should be followed to meet these goals?
   (Phase 2 - Modules 6-7)
3. Do these strategies work, and what changes are needed?
   (Phase 3 - Modules 8-10)
4. Did the program meet its goals?
   (Phase 4 - Modules 11-12)

This module directs itself to the last question -- did the program meet its goals? It tries to impart the skills needed to obtain an answer, the skills involved in planning and implementing a summative evaluation of the end products of the earlier program development phases.

A Model for Developing

Comprehensive Career Guidance Programs



Absolute decision - A decision regarding an educational program involving examination of obtained results in view of the value(s) the decision maker places on the results; i.e., decision about the degree to which program goals have been obtained with a given target group.

Coding data - A method of simplifying response data to facilitate scoring and interpretation; for example, labeling response options in a multiple-choice item "a," "b," "c," etc.

Comparison group - Any group not exposed to the evaluated program(s) used for comparison with the experimental group that is exposed to the program in an evaluation design.

Control group - A group which is chosen by randomization and not exposed to the evaluated program, used for comparison with the experimental group which is exposed to the program.

Cost-effectiveness - Summative evaluation criterion which reflects relationship(s) between program costs and program impact or effectiveness; to assess these relationships requires evaluation data on program outcomes and on cost per program, participant, or objective.

Criterion (criteria) - Concept or set of concepts used as a standard in making judgments. For example, an evaluative criterion can be a statement indicating the standard of performance a student must meet to show he has achieved an objective.

Criterion-referenced test - Instrument which measures the performance of an individual in relation to a predefined performance standard. Criterion-referenced tests are usually based upon the objectives of an instructional program.

Evaluation design - The plan and schedule by which an evaluation is conducted, usually providing for the gathering of baseline (pretest) and post-test data and a control or comparison group.

Experimental design - Category of evaluation design which provides for the collection of pretest and posttest data and for a control group; subjects are randomly assigned to all subgroups.

External validity - The extent to which evaluation results may be generalized to settings and groups other than the ones evaluated.

Field test - Trial administration of a preliminary form of an evaluation instrument to a sample of individuals. Its purpose is to discover problems, prior to large-scale use, in the construction of an instrument, such as ambiguous questions, unclear directions, administrative difficulties, etc.

Formative evaluation - An evaluation process, carried on during the developmental stages of a program, designed to provide information to assist developers to improve and revise the program prior to large-scale implementation.

Guidance - A generic term, including counseling, placement, and follow-through, and intended to refer to the entire program.

Instrument - A data or information-gathering device used in evaluation, such as paper-and-pencil tests, interviews, questionnaires, rating scales, etc.

Internal validity - The extent to which evaluation results may be attributed directly to program effects and not to outside factors such as external events, maturation, inadequacies or changes in evaluation procedures, etc.

Norm-referenced test - Instrument which measures or describes the performance of learners in comparison with the performance of some previously-tested group, such as "third graders throughout the U.S."

Quasi-experimental design - Category of evaluation design which provides a comparison group but allows non-random selection of subjects to all subgroups.

Random selection/sampling - A method of selecting individuals for evaluation subgroups so that each individual in the population has an equal chance of being selected; aimed to avoid bias in the composition of groups.

Relative decision - Decision regarding an educational program involving examination of obtained results in comparison with results attained by other possible programs having similar objectives and target populations; i.e., decision about the potential effects of a program on similar target groups.

Sample - A selected portion of a total population or universe.

Statistical summary - A means of summarizing data entailing use of mathematical or statistical procedures, such as calculating percentages or measures of central tendency (averages).

Summative evaluation - An evaluation process carried on after a program has been completely developed and implemented, designed to provide information on the overall effectiveness of the program to assist decision makers to determine the future course of the program.

Test of statistical significance - A statistical procedure designed to determine the extent to which results (evaluation) differ from chance expectations, or the likelihood that results were caused by program differences and not chance or sampling error.

Unobtrusive measure - A means of obtaining data without the individual being observed or being aware that data are being recorded for evaluation purposes; for example, recording the types of library books being checked out to determine the nature of students' reading interests.

DESIGNING A SUMMATIVE EVALUATION

- *A. Rationale*
- *B. Design concerns*
  - *1. Internal validity*
  - *2. External validity*
- *C. Specify Resources*
- *D. Specify Target Population*
- *E. Select Outcomes to be Evaluated*
- *F. Select Data Summary and Analysis Methods*
- *G. Types of Design*
- *H. Sampling Considerations*

RATIONALE

Why Evaluate

To Provide Concrete and Timely Information for Decisions

Throughout the process of developing a comprehensive career guidance program, evaluation and decision making take place. Once a program has been implemented, however, the need for evaluation and decision making takes on a new purpose. Summative evaluation is required at this point to provide information for major program decisions by teachers, counselors, administrators, even taxpayers. How is it different?

Accountability

Summative evaluation provides for program *accountability*. Accountability applies measurement techniques to answer questions related to ultimate measurement techniques to answer questions related to ultimate program impact. Was the program successful? Should it be continued? In what major ways should it be changed? These are the kinds of questions to be answered at this stage.

Program Improvement

In addition, the *process* of summative evaluation contributes to motivation for program improvement. In this sense it is still formative. Indeed, many experts contend

all evaluation is formative, as it all contributes to program improvement. The difference is that here major decisions are involved about continuing or ending whole segments of programs, as opposed to smaller fine tuning efforts earlier.

When to Evaluate?

When to Evaluate

The timing of a summative evaluation depends upon its purposes. When program personnel need information in order to make major decisions, an evaluation is desirable. However, it is dangerous to attempt a summative evaluation near the beginning of a project, in its early implementation stage. Although you can begin planning for the evaluation early, it is desirable to wait until the program is underway before you begin the summative assessment; in some cases this may be at the start of its second year.

Wait Until Program Is Underway

What to Do

Identify Decisions To Be Made

To aid decision making, anyone who conducts summative evaluations must first identify the decision makers who need information. Then the evaluators must help the decision makers *specify the decisions for which they need information.* The following questions regarding a program's outcomes and its future are typical.

- Is the program achieving its goals and objectives?
- Is the program producing any unanticipated side effects, either positive or negative?
- What does the program cost over and above regular expenditures or compared with other programs, and do the outcomes justify the cost?
- Should the program itself be continued and, if so, what modifications would facilitate even more positive program outcomes?

DESIGNING THE EVALUATION

When an evaluation is carefully designed, its results can be interpreted easily and often can be generalized. The design seeks to ensure that evaluation finds are attributable to the program and not to other extraneous factors. The evaluation design helps to ensure *internal validity*[1] of results by reducing the occurrence of outside factors, sometimes called *threats to internal validity*. These factors include such things as outside events (e.g., a T.V. course on career planning similar to the guidance program), natural growth and maturation, effects of initial testing on subsequent testing, and preexisting differences among groups in abilities and motivation.

Internal Validity

Threats to Internal Validity

The following case study may give you a better understanding of the problem of evaluation design and internal validity.

Illustration

> Lehigh High School wanted to evaluate the effectiveness of a new four-week drug education program using a text developed specifically for the program. The evaluation design was to administer a test to the students before the program began and at its conclusion.

Although the pre- and posttests allow measurement of gains, do they insure that measured gains are outcomes of the program? What are some possible alternative explanations -- threats to internal validity? Suppose that concurrent to the drug education program at school, a similar T.V. program was aired. Suppose also that taking the pretest had an effect on what students were willing to learn or on their familiarity with the procedures for taking tests. Both factors would likely improve student scores

[1]This use of the term *validity* should not be confused with the use which refers to a desirable property of measurement techniques. See Module 10 for an explanation of this property, as well as a discussion of other criteria of measurement adequacy.

regardless of effects of the course. Suppose also that scores changes could result from inadequacies in the measure or in administration procedures.

Alternative explanations for score changes such as these must be empirically and logically eliminated if evaluation results are to be interpreted clearly. Many of these validity "threats" are most typically handled by establishing a comparison, or control group, and by random sampling techniques which will be discussed below.

When "samples" of students take part in an evaluation, you may become concerned about generalizing their results to other similar groups. The extent to which results may be generalized is part of *external validity*. Possible threats to external validity exist and must be minimized by the evaluation design. These questions must be answered: Are students in the study similar to persons to whom results will be generalized? Is the setting for the study free from artificial or reactive effects which affect the way participants respond? Does the pretest influence the post-test? If the results were based on several training processes, are the results relevant to situations where only

External Validity

portions of the program are implemented? It is important to note that although external validity is important, questions of internal validity are of much greater concern in most summative evaluation designs, particularly those relevant to *local* decision making.

Specify Resources

Specify Resources

You need to know the resources available to conduct your study so that any practical limitations imposed by them can be included in your planning. Such resources include

- money
- professional staff time
- clerical help
- student help
- supplies
- computer facilities
- others

Specify Target Population

Specify Target Population

The goals of a comprehensive guidance program are likely to affect a range of persons in addition to students -- teachers, counselors, administrators, and parents, for example. In a comprehensive summative evaluation, you might attempt to measure outcomes for all these groups. It is likely, however, that resources will limit the scope of your investigation and will suggest choosing samples from the above groups.

Select Outcomes To Be Evaluated

Select Outcomes

It may not be feasible to measure attainment of all the outcomes of your program. Your choice of outcomes to measure should depend mainly on the decisions which need to be made, but you may also follow certain guidelines. You might choose to measure achievement of outcomes judged as particularly important by the decision makers. Or you might

select a small segment of a more comprehensive program, such as providing career information or helping clients to know themselves -- one about which there is particular interest. In any event, try to avoid measuring only those outcomes which are easiest to measure. It is better to have limited data on important outcomes than to have lots of data on unimportant ones.

Identify Possible Side Effects

In addition to selecting the outcomes for measurement, you should identify possible positive or negative side effects of the program which should be measured. Here, too, while your focus may be on important student outcomes, it is useful to gather data from other groups to measure any unexpected effects on them.

Select Data Summary and Analysis Methods

Planning for Use of Data in Decision Making

From the start, you should be sure that the information you gather can be used in the decisions that must be made. Your first task is to look at results in terms of the value the decision maker places on them. Decisions based on the perceived value of a program are *absolute* decisions. Absolute decisions ask the question, "Is it worth doing?" The second task is to look at results in comparison to the results attained by other educational programs having similar objectives and clients. Decisions based on a comparison of program results are called *relative* decisions. Relative decisions ask the question, "What is the best way to do it?"

Absolute Decisions

Relative Decisions

While the idea of planning the data summary and analysis for these purposes may seem obvious, it is astonishing how often this is neglected. Many program administrators have put forth substantial efforts in constructive directions, only to find out that when all the data were in, they couldn't be summarized in the way planned, or they lacked some bit of structuring to permit the intended analysis.

Computer Concerns

It is likely that you will use a computer in the summary and analysis of data, and while computers can do marvelous things, they have to be programmed first. This can be tricky business. If you plan to use a computer, consult a programmer at this design stage, make clear just what decisions you want to be able to make, and design the data summary and analysis accordingly.

Types of Designs

Three basic types of designs are available for a summative evaluation: *experimental, quasi-experimental and descriptive designs.*

Experimental, Quasi-experimental Designs

Both experimental and quasi-experimental have provisions for establishing baseline data, usually through a pretest of some kind, and establishing a comparison or control group unaffected by the program. The two design types differ mainly in that subjects are *randomly* assigned to alternative treatments in experimental designs and nonrandomly assigned in quasi-experimental designs. The more bias that can be eliminated from pre-experimental group comparisons through random sampling, the more likely post-experimental comparisons are to produce meaningful estimates of relative impact. Experimental and quasi-experimental designs are discussed in full in Campbell and Stanley's 1966 monograph.[2]

Random v. Non-random

Descriptive Designs

The descriptive design is classified as pre-experimental by some writers. A design which lacks extensive controls for external effects, it is used when the more rigorous designs are impractical. Even with this design, it is important to collect as much information on program results as possible. Too often educators become discouraged by the practical constraints of a school setting and give up altogether. Do not let the impossibility of a strict experimental design dim your resolve to evaluate. A well-

---

[2]Donald T. Campbell and Julian Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching," in *Handbook of Research on Teaching*, ed. N. L. Gage (Chicago: Rand McNally and Co., 1963), pp. 171-246. Reprinted as *Experimental and Quasi-Experimental Designs for Research*. (Chicago: Rand McNally and Co., 1966).

planned and carefully thought-out evaluation which merely provides a description of the results of the program can be very valuable.

Sampling Considerations

Sampling

When selecting an evaluation design, you often need to define the population and select a sample. When you cannot include everyone, select a sample as representative of the entire population as possible.* Making a random selection of the sample assures that the sample is representative of the population. Random sampling allows each person an equal probability of being included; the selection of one unit has no influence on the selection of any other. The best procedure, which makes use of a table of random numbers, follows:

- Assign each unit of the population a number.
- Select random digits from a table of random numbers, beginning from a place on the table selected by blindly placing the point of a pencil somewhere on the page and taking numbers in some predetermined direction.
- Select for the sample the unit of the population having that number. If the number appears a second time, select another.

Here is a portion of a table of random numbers. Such tables appear in most statistics books.

ROW NUMBER				
1	50691	91653	88574	08675
2	19787	66937	91769	13399
3	16746	77983	18061	23664
4	91039	16099	38824	00778
5	11075	62081	88977	78676
.	.	.	.	.

*Module 3 treats this topic in more depth, if you would like to pursue it.

If you know important characteristics of the population, you can use stratified random sampling. With this method, you identify relevant factors such as age, sex, or grade level and draw random samples from the strata. If you are interested in exploring sampling in depth, see the Reference section and Module 3.

Summary

Summary

When you design a summative evaluation you must specify

- available resources
- target population
- target outcomes

You need an evaluation design to ensure that changes which occur are in fact attributable to the program and not to other factors. In other words, the design strives to ensure internal and external validity of the results.

Three basic designs are possible:

- experimental,
- quasi-experimental
- descriptive

The first two designs provide for baseline data and a comparison group but differ in that subjects are randomly assigned to groups in experimental designs and non-randomly assigned in the quasi-experimental designs. The third design includes less rigorous approaches that are resorted to because of practical constraints.

Finally, you should develop a sampling strategy, where needed, and keep biasing factors to a minimum through stratified random sampling when appropriate.

DISCUSSION

UNDERSTANDING THE PURPOSES AND VALUE OF CONDUCTING SUMMATIVE EVALUATION

The purpose of this discussion is to help you develop an understanding of the need for and potential benefit of conducting a summative evaluation in your own school setting.

Break into small discussion groups (three to five persons). If possible, get together with persons who have the same role in the school system -- i.e., all district-level administrators group together, school administrators, counselors, teachers, and so on. Then briefly discuss each of the following:

1. Major decisions you've made over the last six months to a year and how you arrived at them
2. How the decision-making processes used might have been more effective. (Could you have used more concrete data?)
3. Kinds of accountability demands you presently face
4. Types of decisions you need to make and/or questions you want to answer over the next six months to a year

Keep your discussion of these points in mind as you continue work on the module, particularly as you plan your own evaluation in the Application portion of the module.

ACTIVITY 1 - DESIGNING A SUMMATIVE EVALUATION

Imagine that you are the Director of Guidance Services for the Lehigh School District, a suburban district made up of ten elementary schools, five junior high schools, and three high schools. Your district is an unusually well-funded one, supported by a fairly homogenous upper-middle class constitutency that is closely involved with the schools. Consequently, substantial resources over the past years have been devoted to developing a comprehensive career guidance program for your students, grades K-12. As a result of assessing student needs and current programs, guidance objectives and a core educational program were developed. This program consists of a series of career guidance learning units which focus on three major goal areas: academic learning, educational/vocational, and inter-/intra-personal goals. The learning emphasis in all units is on the acquisition of knowledge and attitudes which will facilitate the attainment of broader program goals.

All units (approximately three per grade level) have been intensively pilot tested during the last two years at two elementary schools and at one junior and one senior high school, with very promising results. Final revisions have been made, and you plan to implement the program district-wide next year, hoping to expose all students in the district to the entire core program. The program will be integrated into the social studies curricula at all grade levels and will be taught by teachers and counselors.

Of course, you want to know how effective the program will be in meeting its multiple objectives. Although you've had ample funds in the past for program development, your budget for evaluation is more limited. You can, however, count on substantial assistance from the teaching and counseling staffs who have implemented the program.

Your task now is to design an appropriate program evaluation, taking into account the variables described above.

Get together with one or two other "Directors of Guidance" to plan your evaluation. Discuss the following questions, and write down (briefly) what <u>you</u> think are the best answers for each.

1. Do you need a design for this evaluation? Why or why not?
2. What resources can you use?
3. What will be the target population involved?
4. What objectives will you measure?
5. What is the <u>best</u> design? Will practical constraints in this situation make this impossible? What one would you choose?
6. Will a sampling strategy be necessary? What or why not? How would you go about selecting a sample?

SELECTING/DEVELOPING THE INSTRUMENTS

A. *Identifying appropriate measurement techniques*
   1. *Norm-referenced measurement*
   2. *Criterion-referenced measurement*
   3. *Various types of instruments*
   4. *Use of multiple measures*
   5. *Unobtrusive measures*

B. *Developing measures*
   1. *Using existing instruments*
   2. *Designing new measures*

Once you have established your evaluation design, it is time to consider just what approaches you will use to measure program outcomes. You can use two basic varieties of tests: norm-referenced and criterion-referenced measures. In evaluating a guidance program, you may use standardized norm-referenced tests, which show how participating students fare as compared with the population used for establishing the norms. This kind of instrument can be useful if the criterion of success for an outcome has been defined in terms of how well your students do compared to a group not in your program. If, on the other hand, you have determined a criterion of expected performance for an outcome, criterion-referenced tests, which are designed to produce data that focus on what a person knows or can do rather than how s/he compares with others, would be your choice. Criterion-referenced tests are appropriate for evaluation purposes which require determination of whether or not students have met the outcome criterion.

Norm-referenced Measurement

Criterion-referenced Measurement

Measurement Techniques

A variety of measurement techniques within these categories, and beyond the use of tests, can be used: paper-and-pencil instruments including short-answer and essay tests, checklists and rating scales, questionnaires, and inventories; performance tests; interviews; and behavioral observations. The type of measure used is determined by the performance outcomes being measured. For example, cognitive outcomes can be measured by short answer and essay tests; attitudes can be measured by questionnaires, inventories, or rating scales; behavioral outcomes can be measured by observations, preferably as a function of units of time. Whenever possible, use several measures, since any

Technique Depends on Type of Outcome Being Measured

Multiple Measures Desired

Behavioral Observation

single measure is only one approximation of the outcomes being measured. Evaluation experts today tend to recommend multiple measurement because effects demonstrated by a number of different techniques are likely to produce greater confidence in the underlying reality of the observed effect.

Unobtrusive Measures

Unobtrusive measures, such as attendance, use of materials, grade point average, and referrals to the counseling

office, can be useful in estimating unanticipated side effects or outcomes. Data of this kind can serve as another index of program success. An advantage of incidental or unobtrusive measures is that the data-collection procedures do not influence the results because the persons are unaware that data are being collected.

Developing Measures

Developing Measures

Developing appropriate measures is demanding. Before deciding to use already existing instruments, you should make a careful review of the items. The publisher's description usually is not adequate to tell you if the instrument is appropriate for the measurement of your program outcomes. Therefore, look carefully at the test itself to discern the purpose and content. Include in this scrutiny the kind of population the test was normed on, especially in culturally distinct communities

In addition to purpose, content, and norming population, there are four other important considerations:

- Time needed for administration
- Ease of administration
- Ease of scoring
- Ease of interpretation

Sources of Test Information

In a search for already existing measures, you can refer to a number of sources to answer different kinds of questions. Thorndike and Hagen (1969) have summarized the sources and types of information on a variety of instruments (See Table 1 below.)

Table 1

SOURCES AND TYPES OF INFORMATION ON INSTRUMENTS

Sources	To Find Out, in Any Field					
	What tests there are	What new tests there are	What test X is like	What specialists think of test X	What facts we have about test X	What research has been done on or with testing problem Y
Texts in special areas of measurement	•		•	?	•	•
Mental Measurements Yearbook	••	?	•	••	•	••
Reviews in current professional journals		••	•	•		
Publishers' catalogues	•	••	•			
Test blank and manual			••		••	•
Review articles in Review of Educ. Research, etc.		•		?	?	•
Tests in Print	••					
Psychological Abstracts		•			•	••
Education Index		•				•

Key ••Most helpful •Somewhat helpful ?Possibly helpful[3]

[3]Robert L. Throndike and Elizabeth Hagen *Measurement and Evaluation in Psychology and Education,* 3rd ed. (New York: Wiley & Sons, Inc., 1969), p. 253.

A good place to start with this investigation is the *Seventh Mental Measurements Yearbook* (Buros, 1972). Those tests not reviewed in earlier editions (and those previously reviewed that have been revised) are described and criticized by educational and psychological authorities.

Buros

Tests in Print (1974) is a comprehensive test bibliography and index to the first five books in the Mental Measurements Yearbook series. Each test mentioned in *Tests in Print* includes the following information:

Tests in Print

- Test title
- Appropriate grade levels
- Publication date
- Special short comments about the test
- Number and type of scores provided
- Authors
- Publisher
- Reference to test reviews in *Mental Measurements Yearbooks*.

Test publishers' catalogs are a particularly good source for locating new and recently revised tests. These catalogs provide basic information about the purpose and content of the test, appropriate level, working time, cost, and scoring services available. An important piece of information that is not provided in all publishers' catalogs is the copyright date (or norm date). A selected list of publishers' names and addresses is included in the Appendix.

Catalogs

Reviews

Other sources of information are the test reviews found in the professional periodicals. Journals such as *Educational and Psychological Measurement*, the *Journal of Educational Measurement*, the *Journal of Counseling Psychology*, and the *Personnel and Guidance Journal* typically carry reviews of some of the more recently published or revised tests. *Educational and Psychological Measurement* also publishes a validity studies section twice a year. The bulk of the articles in this section are reports of studies using various standardized instruments for predictive pur-

poses. Textbooks on measurement also typically include information on various tests.

Need to Develop New Measures

Although it is obviously easier and cheaper to use or modify an existing instrument, the advantages quickly evaporate if it does not fit the outcomes you want to measure. Having readily interpretable data may justify the extra effort needed to design new measures. It is important to note, however, that if you follow this path, you must take care to insure adequate validity and reliability.*

Developing and Refining the Measures

Develop Measures

If you develop new instruments or adapt existing ones, take these three steps:

1. Develop items which fit outcomes and the type of measurement technique(s) selected.
2. Pilot test the drafted items on a small group of students similar in characteristics to the target group.
3. Make revisions and refinements based on tryouts prior to the actual evaluation.

If you choose this route, study Module 3 which treats the subject of developing instruments in some detail.

Summary

Summary

The major steps to follow in developing measurement instruments are

- Select type(s) of instrument needed to measure program outcomes.
- Examine existing instruments.
- Decide to adopt, adapt, or create.
- Adapt or create measures and data-collection form, if necessary.
- Pilot test and revise.

*Module 10, Trying Out Activities and Monitoring Early Implementation Efforts, contains background information on measurement techniques and criteria of adequacy (validity, reliability, objectivity and efficiency).

ACTIVITY 2 - SELECTING/DEVELOPING EVALUATION INSTRUMENTS

The purpose of this activity is to help you develop some of the basic skills needed in selecting and developing appropriate evaluation instruments. First look at the examples of outcomes and assessment techniques on the next few pages. There are only a few examples from many possibilities. Also review the discussion of possible measurement techniques presented in the text.

Assume that the behavioral objectives below are a part of your new, comprehensive guidance program and that they will be measured as part of your summative evaluation.

a. For each objective, name an appropriate measurement technique.

b. For each objective develop one item which measures achievement of that objective. You might want to use a chart which looks something like the following:

Objective	Appropriate Measurement Technique	Sample Item
1		
2		
3		
4		
5		

Objectives

1. The student will list three reasons why it is important for a person to consider several educational and vocational alternatives in the process of setting long-range goals.
2. The student will identify her/his abilities, interests, and values.
3. The student will indicate that s/he is willing to seek help from other persons in solving a variety of personal problems.
4. The student will demonstrate an effort to find vocational and educational information by checking out at least one item per month from the career guidance library.
5. Given hypothetical situations, the student will identify how s/he would feel in the situation.

Examples Of Outcomes And Assessment Techniques

The first example below shows a performance objective and one of the multiple-choice items written to measure its accomplishment. This is a closed form of questionnaire.

Performance Objective: Given examples of factors which are affecting job opportunities, identify whether they are social, economic, or technological factors.

Test Item:

When a new industrial process is invented and is introduced into an industry, sometimes certain jobs are no longer needed. In other cases, however, increased efficiency in the industry creates many new jobs. These are examples of a

a. trend
b. social factor
c. economic factor
d. technological factor

The second example presents another performance objective but, this time, one which uses a more demanding, open-ended questionnaire format.

Performance Objective: To make an informed selection of your first-choice tentative long-range vocational goal and of your second-choice long-range goal.

Questionnaire Item:

1. During the last quarter, did you pick some tentative long-range vocational goals? Answer this question by putting a check in the appropriate box below.

☐	☐
Yes, I picked a first- and second-choice vocational long-range goal.	No, I did not pick a first- and second-choice vocational long-range goal.

If you checked the "Yes" box, there are a few additional questions for you to answer below and on the following pages. If you checked the "No" box, you are finished with this survey and may hand it in to your counselor.

2. In the spaces provided below, describe in a statement or two the occupational families which you selected as your two long-range goals.

3. In the charts provided, indicate how your developed abilities compare with the abilities of twelfth graders who later entered occupations related to your goals.

4. In the spaces provided, describe how your interests are related to your first and second-choice goals. Mention the results of any interest inventories you may have taken during the last quarter.

5. In the spaces provided, describe how you considered your work values before you picked your first- and second-choice goals. Mention the results of any work values inventories you may have taken during the last quarter.

6. In the spaces provided, describe how you considered your personal and social behavior before you picked your first- and second-choice goals.

7. In the spaces provided, describe how you considered your physical traits before you picked your first- and second-choice goals.

The third example includes a performance objective that specifies an attitude outcome and an item taken from an attitude survey.

Performance Objective: Students will identify as their preferred method of choosing career goals that of having advice from someone else but selecting the goals themselves.

Survey Item:

Each phrase at the left is the beginning of a sentence. From the statements at the right, select the one that would make the sentence true for you. Then place the letter corresponding to that statement on the line next to the phrase.

In regard to:

my educational goals, I prefer to (put one of the letters at the right here) _____

my occupational goals, I prefer to _____

my social-relationship goals, I prefer _____

my personal-behavior goals, I prefer to _____

A. have these goals selected for me by someone else.

B. have someone else choose goals for me.

C. have advice from someone else but select the goals myself.

D. select these specific goals myself without help from anyone else.

The fourth and final example illustrates a student behavior observation procedure that has been used to evaluate changes students have made in employing certain personal and social skills.

Performance Objective: To increase the number of positive behaviors students perform in the area of showing consideration for the feelings of others.

Behavioral Observation by Teachers:

A Performance Record is used by teachers to help students gather facts about their behavior. The Performance Record consists of a list of behaviors which are important incidents in the personal and social development of students. For the above objective, the following student behaviors are listed on the Performance Record:

Shared or took turns with someone.
Helped to assure that others were not left out.
Helped another person to participate in learning or play activity.
Aided another person who was not feeling well or who was physically hurt.
Helped a handicapped student to participate in learning or play activities.
Provided something for or gave up something to help another person.
Cheered up, complimented, or encouraged another person.
Showed respect in making suggestions to another person.
On his own initiative, interceded for another person who was being criticized or taken advantage of by others.

Teachers are encouraged to record the critical incidents as objectively as possible. The recording is considered objective if another teacher, knowing a student as well and observing the student at the same time, then reports the same critical incident. Teachers should not set aside a definite time each day to record the critical incidents of student behavior. Nor should they observe a student with the hope he will do something which could be recorded as a critical incident. Instead, teachers should record the critical incidents of student behavior which are noticed spontaneously during the course of the school day. Furthermore, behaviors should be continual. This does not mean that teachers will have to spend a great deal of time observing; minute-by-minute observation is not necessary. Only when noteworthy behaviors are performed do they need to take notice for the purpose of recording critical incidents.

ADMINISTERING THE INSTRUMENTS

A. *Scheduling the administration*
B. *Selecting and training administrators*
C. *Orienting students*

Once the instruments are developed, a number of tasks must be done as part of the data-collection effort.

Scheduling the Administration

Schedule Data Collection

During your preparatory activities, you should write a detailed schedule for data collection. You will need to decide how students should be grouped for testing. Will whole classes be tested? Small groups? Individuals? How many times? Your schedule should include training sessions for administrators, orientation sessions for students if necessary, and the testing itself. It should ensure that regular school schedules are disrupted as little as possible. In this regard, it is usually best to keep the schedule fairly compact, collecting all the data in as short a period as practical. Rooms should be set aside which are adequate for the size of the groups involved. You should establish the schedule as early in the evaluation planning as possible and make it available to all participants. In this way, you will have adequate time to make any necessary changes and will be able to coordinate efforts.

Selecting and Training Administrators

Select Administrators

You may get people from a number of sources to administer the instruments. If the evaluation focuses on paper-and-pencil instruments administered in the regular classroom, you can probably use regular classroom teachers. If other types of measures are employed, such as interviews or observations, you will often need outside help. Sources for help include local college students, substitute teachers, paraprofessionals in the school, or community

residents. It is important to fit the type of person to the task; for example, a substitute teacher might be a more appropriate interviewer than a clerical aide, who would be well suited to examine school records or documents. School personnel may ease the data-collection process through their familiarity with school routines, and they may facilitate unobtrusiveness if desired. However, they may affect the results of measures where students wish to make anonymous responses (e.g., in an interview). Also, evaluation by regular school personnel is sometimes subject to criticism in terms of the reactive effects on susceptible measures. Thus you should use school personnel cautiously to minimize possible biases.

Train Administrators

Once you have chosen instrument administrators, you must train them. The amount of training naturally depends on the task to be done. Training is likely to be more effective if it uses a combination of workshop meetings, written material, and practice exercises. The workshop

permits you to acquaint administrators with the testing procedures and relevant information about the school and to distribute written instructions and material. Written instructions should explain in complete detail the process they will go through. When everyone understands the instructions, they can practice by role playing. They should also be informed of the testing schedule -- where and when to appear -- and any necessary changes should be coordinated.

Orienting Students

Orient Students

Students may require some orientation to the instrument. Obviously, if you use only non-reactive instruments such as behavior observations, no orientation is needed since students should not know something special is happening. However, if you use such devices as paper-and-pencil tests, interviews, or attitudinal surveys, you can tell students about the purposes and the procedures. Through orientation you can not only inform students but also promote cooperation. Orientation should be uniform for all students. It can be conducted by the trained test administrators, and is most efficiently held with class-sized groups. During orientation, you can use written materials, such as a letter explaining what student participation will involve, to supplement a verbal presentation. You can also distribute schedules for testing, if needed for individual or group activities that occur outside the regular classroom.

Summary

Summary

To review, the tasks involved in preparing for and collecting data are as follows:

1. Prepare a schedule of preparatory and administration activities.
2. Select and train test administrators.
3. Orient students, if necessary.

One final thought. Module 3 deals somewhat more extensively with the topic of administering instruments, in relationship to conducting a desired outcomes assessment. If you're interested in exploring this topic further you might want to check there.

DISCUSSION - ADMINISTERING THE INSTRUMENT

Think through the following questions in terms of your own setting.

1. What will you have to do to carry out the actual data-collection effort at your school? Consider the tasks involved to accomplish those things listed in the text. What other organizational and preparatory efforts can you think of that will have to be done?

2. Who might be good as data collectors, or administrators, at your school?

3. Have efforts of this sort been done in the past at your school? What problems occurred? How were they dealt with?

4. What expert help in collecting data is available to your program?

PROCESSING AND ANALYZING THE DATA

A. *Processing data*
   1. *Handscoring*
   2. *Computer scoring*
   3. *ADP scoring*
   4. *Tallying frequencies*
   5. *Organizing data by characteristics of subjects*

B. *Analyzing the data*
   1. *Method of data presentation*
   2. *Tests of statistical significance*

Measurement instruments are administered according to an evaluation design in order to make inferences about the effects of the program. The choice of appropriate processing and analysis procedures depends largely upon the items in the instrument and the evaluation design. These, of course, as noted repeatedly, should derive from the decisions to be made.

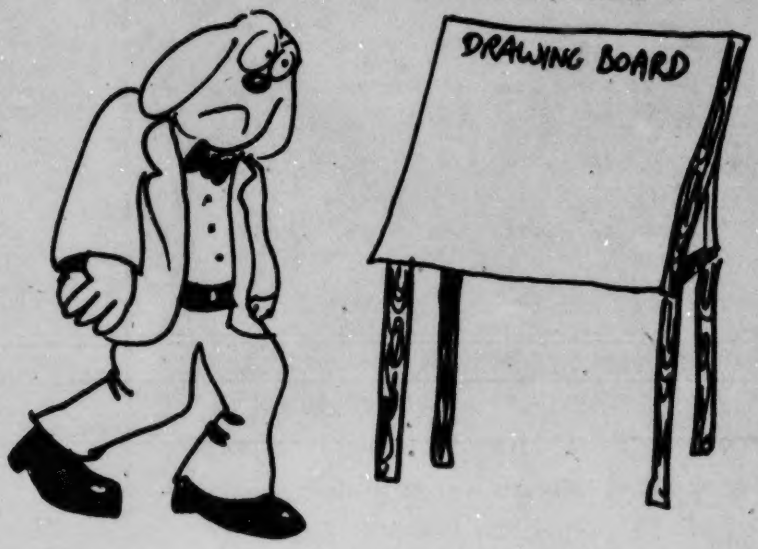
Processing Data

Processing Data

It is possible that much of your data processing will be done manually. This is especially true when samples are small and resources limited. It is good practice to have someone spot-check hand scoring to make sure it is being done accurately. Sometimes hand scoring entails making judgments to determine if a response is adequate or to translate a lengthy response into a discrete score. On items where such judgment is needed, several points should be considered. First, does the scorer understand the criteria and apply them consistently over time? Second, has objectivity been promoted by concealing identifying information from scorers? And third, if two or more scorers divide the task, would they mark the same answer in the

Scoring by Hand

same way? An answer to these questions can be facilitated by having a random sample of student responses scored twice. If the results are not consistent, it would be desirable to rescore the troublesome items after retraining the scorers.



Computer Scoring

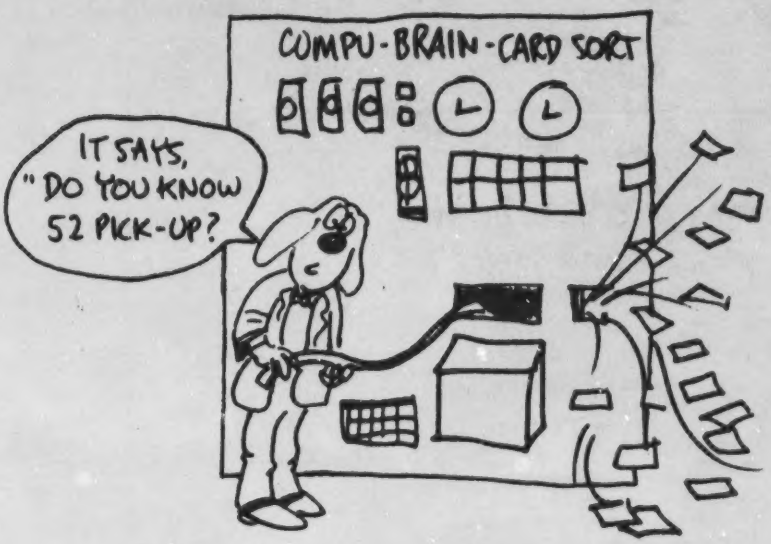
In many cases, the easiest way to simplify the task of processing data is to have a computer do it for you -- if you have access to a computer and resources for it. Your task then involves recording data onto data cards so that the computer can process them and determining what kind of output you want from the computer; i.e., what information you want the computer to print out and in what format you want it printed. A computer programmer can then write or select a program to tell the computer what you want, and the computer will print out your results. You will then have a master list with all your required information organized in the way that you want.

ADP Scoring

Often, however, you will not have access to a computer or you will not be working with large enough samples of data to warrant using a computer. There are other kinds of automatic data processing (ADP) equipment which may be

available. One of these is a card sorter. You record your data on data cards by keypunching, just as you would have done for computer analysis; the sorter will then sort and tabulate these cards numerically or alphabetically.

It is important to note here that the ease and convenience of computer processing vary with the task. Many uninitiated researchers and evaluators, particularly those with limited resources, wind up with colossal headaches because of problems with computers and ADP.



Tallying Frequencies

A common method of processing observation data is to tally or compute frequencies. Suppose that you observe a classroom to determine how well students demonstrate being in touch with their feelings (an objective of your program). You note how often individuals state how they feel about something during a guidance-related discussion. Your raw data might look like this:

Students expressing feeling	Time
Susan	10:12
Tom	10:15
Susan	10:16
Mike	10:23
Tom	10:25
Susan	10:27
Susan	10:35
Tom	10:38
Mike	10:42
Carol	10:46

You may note little problems in the personal and social areas...

After tallying, the data might look like this:

Student	Time Period 10:10-19	10:20-29	10:30-39	10:40-49
Susan	1 1 (2)	1 (1)	1 (1)	
Tom	1 (1)	1 (1)	1 (1)	
Mike		1 (1)		1 (1)
Carol				1 (1)
TOTAL	3	3	2	2

Organizing Da... by Characteristics of Subjects

It is often necessary to organize data, as part of processing it. Many times in evaluation you will be interested in examining the data for specific groups of subjects, such as a particular age or grade level, a particular sex, and so forth. If so, you would want to arrange the data into relevant groups (fourth, fifth, and sixth grades, males and females, etc.). Other times you may simply want to arrange the data in some logical order so that you can find data for particular students more easily. Alphabetizing the data according to the subjects' names is an example of this. Assigning code numbers is often helpful in organizing data, and it helps to conceal individual identities.

Analyzing the Data

Data Analysis

In summative evaluation, the processed data produced by outcome measurement is analyzed to aid decision making. Instruction on sophisticated data analysis is outside the scope of this module. Since many of the evaluations you will be concerned with do not require highly developed analytic skills, we will focus on increasing the skills and procedures useful in analyzing these less sophisticated evaluations.

Methods of Data Presentation

Methods of data presentation. Regardless of the type of decision to be made, you will have to have some type of descriptive data presentation. Absolute (value-oriented) decisions would require only a presentation of the target group's data, while relative (comparison-based) decisions would require data from an additional group. Presentation may be as simple as listing the raw scores generated by the measurement techniques. Often, however, statistical summaries or analyses offer a better basis for decisions.

Tables, Statistical Summari and Graphs

Descriptive data may be presented in *tabular, statistical, or graphic* form. An example should help clarify the differences among these three methods. Suppose that a guidance survey test was used as part of an evaluation to assess achievement of ten performance objectives.

The degree to which students achieved the objectives could be presented in *tabular* form as in Table 2 on the next page. It describes the number of objectives achieved by all students.

Percentages

These numbers are also transformed into the percentages of students achieving each objective. These percentages are a type of *statistical summary* of the data. Proportions and ratios may be similarly employed. Another type of statistical summary entails the use of averages. For example, the average (mean) number of objectives achieved by the students in Table 2 is 6.

Averages

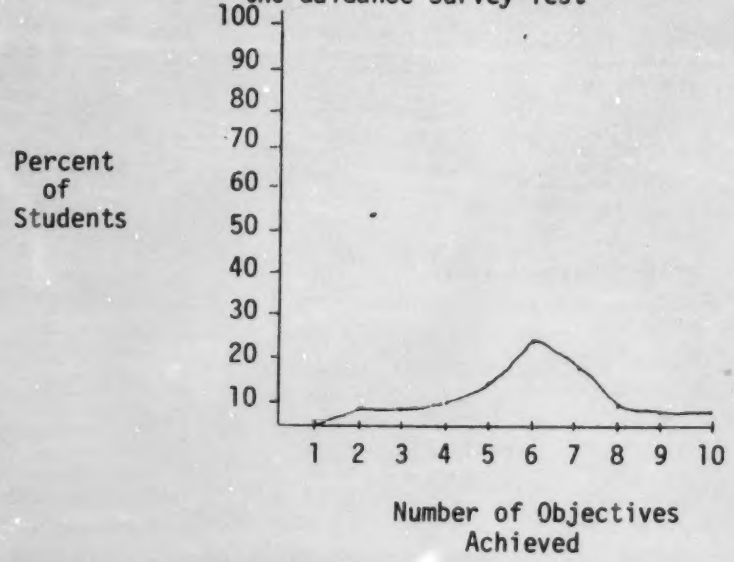
A final method of presentation is the *graph*. The bottom of Table 2 presents the same data in graphic form. Graphs have the advantage of being easy to understand, especially when tabular presentations summarize complex bits of information in ways which are confusing.

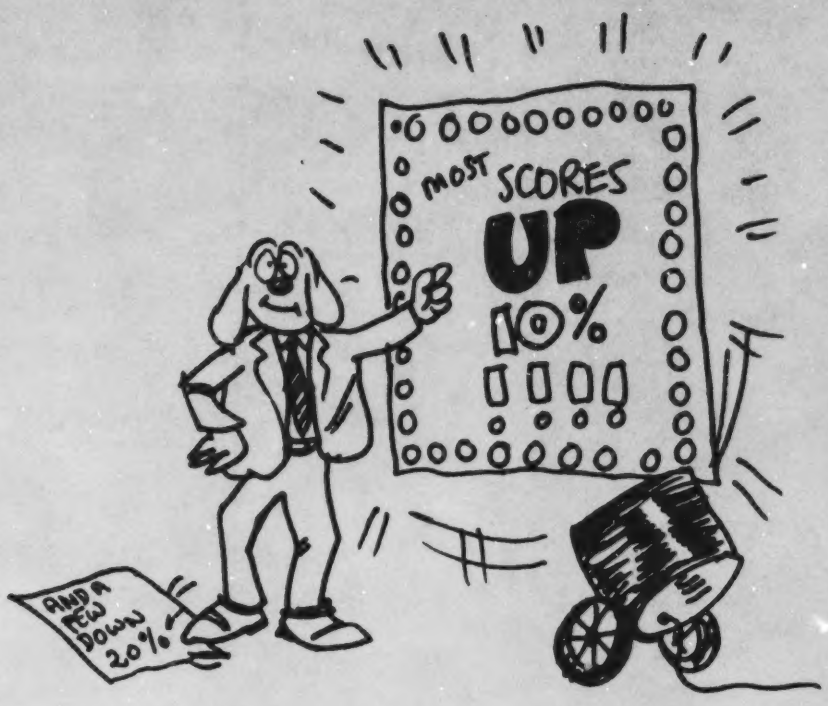
Table 2

Number of Objectives Achieved by Students on Guidance Survey Test

Number of Objectives Achieved	Number of Students	Percent of Students
1	0	0
2	5	5
3	5	5
4	10	10
5	15	15
6	25	25
7	20	20
8	10	10
9	5	5
10	5	5
Totals	100	100%

Graphic representation of the percent of students achieving objectives on the Guidance Survey Test





Tests of Statistical Significance

Tests of statistical significance. For relative decisions, inferential statistics may be used to determine the significance of differences among comparison groups (usually experimental and controls). Tests of statistical significance are designed to determine whether the differences occurring among groups tested are greater than those which would occur by chance alone. We do not have the time or space to make you proficient even in an elementary application of inferential statistics. Thus at this stage of analysis, you may need to choose or hire someone with the expertise to make the more complex analyses needed for a statistical comparison of the evaluation samples.

Summary

Summary

Once you have collected your data, you need to process it. You may process data by hand, computer, ADP equipment, or a combination of these.

After you have processed the data, you must analyze it. This may involve a sophisticated process. You must also present the data. Methods of presentation you can use include tables, statistical summaries, and graphs. To determine if differences between comparison groups are significant, you (or someone who has expertise in inferential statistics) can perform tests of statistical significance.

ACTIVITY 3 - PREPARING BAR AND LINE GRAPHS

This activity will allow you to make some of the summary tables and/or graphs that the evaluator should use.

INSTRUCTIONS:

Using the data in the "Percentage of Responses Data Table" below, please complete the bar graph on the following page which includes the percentage of people in each group who judged the following as excellent:

a. instruction given in the program

b. curriculum materials used in the program

c. communication between personnel working in the project

d. community involvement effort made by the project

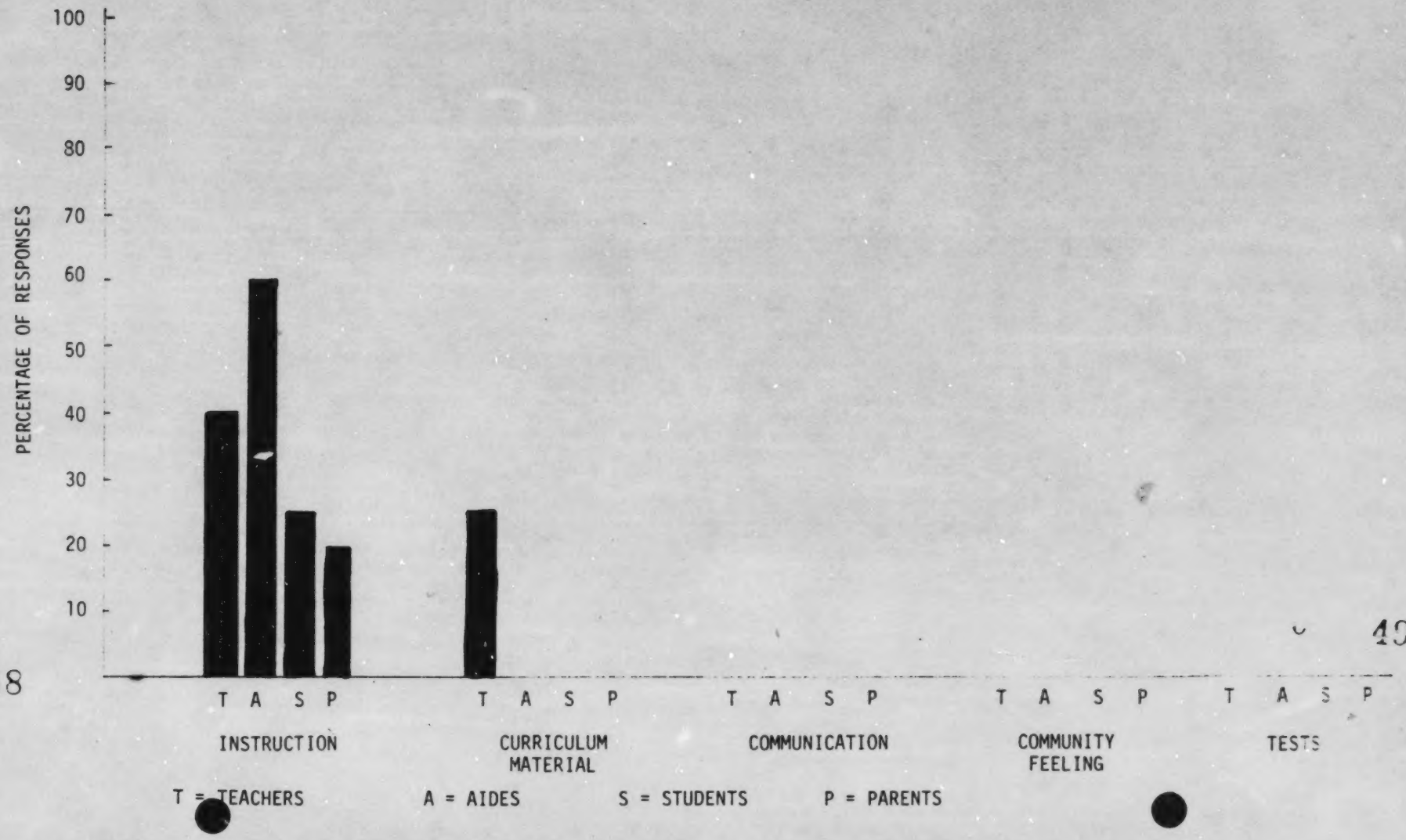
e. tests used to evaluate student's outcomes

PERCENTAGE OF RESPONSES BY QUESTION AND BY GROUP DATA TABLE

Questions		Instruction			Curriculum			Communication			Community Involvement			Evaluation Tests		
Ratings		P	A	E	P	A	E	P	A	E	P	A	E	P	A	E
GROUPS	Teachers	25	35	40	55	20	25	15	35	50	20	10	70	40	40	20
	Aids	5	35	60	10	45	45	65	20	15	15	35	50	10	20	70
	Students	35	40	25	50	40	10	15	60	25	15	70	15	80	15	5
	Parents	10	70	20	30	60	10	35	35	30	20	45	35	25	60	15
ALL		20	45	35	35	40	25	35	40	25	20	40	40	40	30	30

P = POOR A = AVERAGE E = EXCELLENT

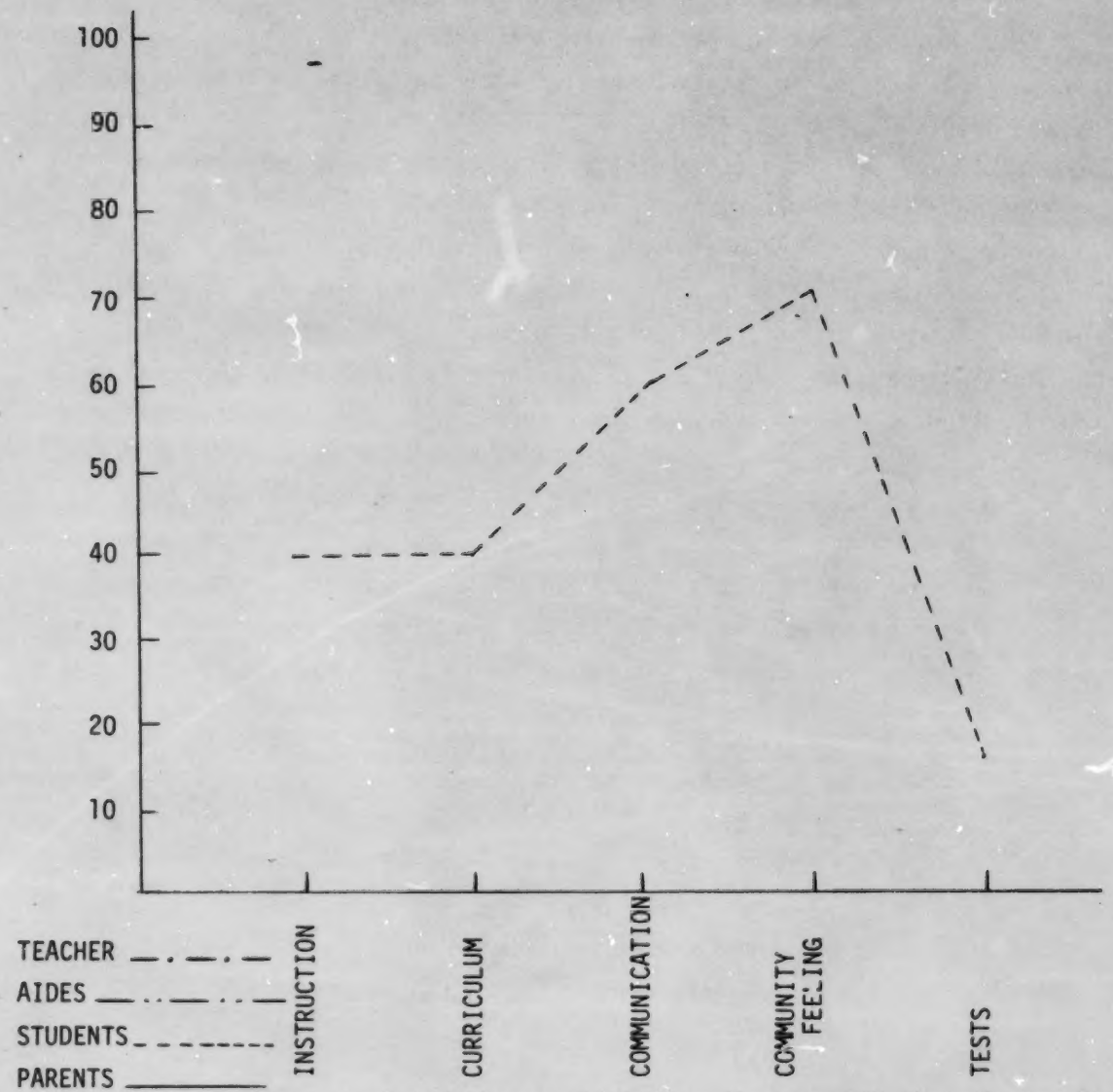
BARGRAPH (HISTOGRAM) REPRESENTATION OF PERCENTAGE OF "EXCELLENT" RESPONSES BY QUESTIONS AND BY GROUPS



Using the data in the Percentage of Responses Data Table, complete a line graph (polygraph) which includes the percentage of people in each group who judged the different components of the program as (a) average. The responses for students have already been graphed to get you started. Graph the responses using the key presented at the bottom of the page.

PERCENTAGE OF "AVERAGE" RESPONSES FOR

PROGRAM COMPONENTS BY GROUPS



When you have completed these activities, discuss with the other members of your group which form is easiest to interpret. Have you seen other methods of data presentation? What form do you like most? Least?

ACTIVITY 4 - PROCESSING AND ANALYZING DATA

Pretend you are still Director of Guidance Services at the Lehigh School District -- congratulations! It is now a year later and you have just conducted the first stage of your summative evaluation. You decided to do it piece by piece, so your first effort focused on junior high school students and two learning units in the inter- intra-personal area. You want to determine what objectives students had mastered before and after exposure to the program and whether differences in growth between program participants and controls are significant.

A survey test covering the objectives of the units was given at the beginning of the semester, and it will be readministered at the semester's end after students have studied the units. At Adams Junior High, where the units were taught, 120 students were pretested, along with a control group of 120 students at Roosevelt Junior High. The pretest data are in and you're ready to process and analyze them. Your resources are limited, but you have adequate clerical staff to assist and access to an educational statistician for consultation. How will you process and analyze the data? (Of course, you planned for all this long ago!)

Your task now is to plan how to process and analyze the data from the pretest and later from the posttest. Get together with two or three other "Directors of Guidance" to plan a strategy, detailing the following:

- the easiest way to process that data
- possible techniques to organize the data
- effective ways to present it

Spend about fifteen minutes in discussion. Then return to your own work place and develop your plan in greater detail. Complete the Planning Form on the next page, and submit your completed forms to the Coordinator as a measure of your achievement of this objective.

PROCESSING AND ANALYSIS PLANNING FORM

Answer the following questions as completely as you can, drawing from your understanding of the module and your discussion with other participants.

1. What is the easiest method for scoring the tests? Why?

2. How might the data be tallied?

3. What are effective techniques for presenting the data?

4. If you wish to compare results between Adams and Roosevelt students once posttest data are in, how will you accomplish this comparison, assuming your own expertise is limited?

RELATING PROGRAM COSTS TO PROGRAM EFFECTS

A. *Determining program costs*
   1. *Realistic estimates*
   2. *Assumptions*
   3. *Chief items*
   4. *Variables*

B. *Establishing cost-effectiveness ratios*

The preceding sections of this module have dealt with gathering and analyzing data on program effects. This section will deal briefly with using such data, along with data on program costs, to make important decisions about program effectiveness and related costs.

Determining Program Costs

Determine Program Costs by Making Realistic Estimates

Module 10 presented suggestions for measuring the cost of activities at the initial tryout stage. During full-scale program implementation, you may find it even harder to determine costs because of greater difficulties in breaking away the costs of one program (personnel, equipment and supplies, and so on) from the costs of others. In general, make the estimate of your program costs as realistic as possible by carrying out the five following steps.

1. Establish discrete input categories, i.e., what people and materials/resources are contributing to the program, and determine the time span you are considering.
2. Determine a basic unit for each category, e.g., person/days, unit costs, etc.
3. Assign a dollar rate or value for each unit.
4. Provide the best possible estimate, supported by program records if possible, of the quantity of units expended during the time span being considered.
5. Compute category totals and overall program cost by multiplying unit quantity times rate and summing.

Table 3

Monthly Costs for Career Guidance Program

Expense Category	No. of Units	Rate	Estimated Total Cost
1. Personnel			
A. Administrators	2 days	$100/day	$ 200
B. Counselors	15 days	$ 60/day	$ 900
C. Teachers	30 days	$ 50/day	$1,500
etc.			
2. Equipment and Supplies			
A. Film rental	1 item	$150	$ 150
B. Textbooks	10 items	$ 8	$ 80
C. Tape recorder and headset	2 items	$150	$ 300
3. Other Program Expenses	--	--	--
4. Building Costs (prorated)	30 days	$ 50/day	$1,500
5. Fixed Costs (prorated)	30 days	$ 17/day	$ 510
6. Total Costs			$5,140

State Assumptions Used in Determining Rates

It is most important that you state the assumptions used in arriving at rates so that users of these estimates can arrive at their own judgments about the rates' accuracy.

Anticipate Changes in Rates

In addition, it is often necessary to anticipate changes in rates since they may change between the time they are calculated and the time the actual program costs can be calculated. This is true of such items as sliding salary scales, administrative costs, building maintenance, and so on.

Chief Items in Determining Costs

Many school programs are likely to have the same overhead costs and thus this variable often need not be computed in determining costs. The chief elements in determining costs are the number of person-hours required, and the

materials and equipment, and facilities necessary (are they currently available? Can they be adapted? Must they be purchased?)

Variables in Relating Costs to Effects

Costs alone tell only half the story. These must be related to the effectiveness and efficiency of the program. Sometimes additional costs are warranted when they produce outstanding results. But even an effective program can have drawbacks if it produces negative unanticipated side effects or if it is inefficient in requiring too much time from the point of view of either students or staff. Thus many variables have to be considered and weighed against each other in relating program costs to program effects.

Establishing Cost-Effectiveness Ratios

Cost Effectiveness Ratios

Much of the literature on cost-effectiveness of cost-benefit[4] analysis presents very complex techniques which are impractical for local applications. The simple cost-

---

[4] *Benefits* in this context usually refer to long-range economic and social gains resulting from improved career guidance programs. Their measurement usually requires sophisticated value estimation procedures which are beyond the scope of most summative evaluation efforts.

effectiveness ratio *is* useful in making decisions about the value of educational programs. The general rule is to use some standard measure of services rendered, preferably outcomes achieved (such as student performance objectives mastered, attitude or knowledge gains demonstrated as the numerator and the total cost estimate as the denominator.

$$\text{Cost-effectiveness ratio} = \frac{\text{Outcome achieved}}{\text{Total cost}}$$

Be careful to use the same numerator and denominator terms for all programs being compared.

Summary

In this section of the module, we have discussed the problem of making estimates of program costs and using them to produce cost-effectiveness ratios. Needs for cost data will depend on your specific evaluation situation. For making absolute decisions, simple program cost data may suffice in the absence of competing programs. On the other hand, if there are alternative strategies for providing the same services or attaining the same outcomes, you may find it very useful to provide cost-effectiveness ratios.

Summary

ACTIVITY 5 - DETERMINING COST EFFECTIVENESS

In this activity you will read about three schools, each of which was developing an exploratory work or leisure experience program for students. The guidance team at each school sought to develop a monitoring system for the exploratory experience program which would ensure that each student was reached. They wanted the monitoring system effective in getting the job done from the standpoint of students and staff and to ensure that students did participate in the program. They also wanted the system to be efficient in terms of the system, staff, and student costs and efforts. Last, they wanted a system which would produce few undesirable side effects.

The systems designed by each of the schools are presented below. They are all evaluating their systems after three months. Read each one and then use the table which follows to rate each of the described monitoring systems in terms of its (1) effectiveness, (2) the unanticipated side effects which it causes, and (3) its costs.

School A

The six members of the career guidance team sent all of the students in the school a notice telling them who their individual development "supervisors" were. The cost for printing all these notices was $25.00 and about three hours of a counselor's time was involved at a rate of $7.00/hr. The students were also told to contact their career guidance team supervisor and arrange to meet with her/him in person after they had had any type of exploratory work or leisure experience.

The career guidance team printed special forms at a cost of $375.00, attractively designed to be appealing to students. The students were each given a large stack of these forms and told to put one in their "supervisor's" box every week, indicating whether or not they were engaged in an exploratory work or leisure experience or what plans, if any, they were making for such experiences. The idea behind this sytem was that having to turn in the forms each week would remind the students to make actual plans for their exploratory experiences and engage in them.

As a result of this system, some students became particularly motivated, so much so that they began reminding their friends each week about the forms and urging them to get actively involved in the program. During the three-month time period covered by the evaluation, 23% of the students turned in at least one form.

School B

The five members of the career guidance team got together and decided that each of them knew a few students well and that each would informally assume responsibility for the students with whom they had good rapport. They decided that the whole thing could be very casual and that there was no need to really formalize the system.

Every two weeks each member of the career guidance team contacted a subset of her/his "counselees." The team members each had a little booklet in which they kept track of which students they had contacted and when. They consulted this booklet for determining when they should next follow up with each of their counselees. This activity took approximately two hours of each guidance team member's time, every two weeks - 12 hours in total for each member. For the three month period of the evaluation, the dollar rate per hour was $7.00.

The career guidance team members found that as a result of the individual development strategy they became even closer to the students, since the system involved them in discussing real-life experiences rather than general, distant plans. This increased closeness also carried over into areas other than work- and leisure-related concerns. The guidance team members reported that they were able to make contact with all of their "supervisees." They are trying to determine cost-benefits after three months.

School C

The three members of the career guidance team got a roster of all the students in the school and used this master-list to divide the students alphabetically into groups of equal size, one group for each team member. This assured that each student would be "covered" by the individual development strategy. The process took one hour of two team member's time.

Once a week, each member of the career guidance team contacted each of the students in her/his group to see if they had had any exploratory work or leisure experiences, and if so, to discuss them. This contact was routinely made on the same day each week, during the student's home room period. This meeting took about three hours of the team members' time each week. Their rate is $7.00 per hour.

This system received criticism from several of the home room teachers who indicated that they felt too much time was being taken away from regular home room activities for this contact. The system was effective in reaching all students. They are trying to compute the cost-benefit ratio after three months.

RATING TABLE

	School A	School B	School C
Effectiveness			
Unanticipated Side Effects			
Estimated Costs (input categories x unit costs)			

After completing your ratings, discuss as a group the monitoring systems developed by each of the three schools and your ratings of each. Which system do you think was the most cost effective?

MAKING DECISIONS BASED ON RESULTS

A. Kinds of decisions

B. Setting criteria for program effectiveness (absolute decisions)

C. Weighing results for relative decisions

1. Statistical vs. educational significance

2. Generalizations for weighing data

Kinds of Decisions

Kinds of Decisions

Having completed a review of the steps involved in conducting a summative evaluation, we come full circle to the original purpose of evaluation, discussed in the first pages of this text -- that is, to facilitate major decisions about a program. The kinds of decisions to be made are specified early in summative evaluation planning, since they determine the course the evaluation will follow. These decisions include:

- continuing or discontinuing the program
- altering the program in basic ways
- expanding the program
- establishing follow-up programs
- modifying or expanding resources
- modifying existing staff and/or staff qualifications

Decisions like these are easier to make when you know if students have achieved program objectives, if unintended outcomes occurred, and how the outcomes relate to costs. When you can pinpoint the objectives the program best helps students achieve and can get an idea of the relative effectiveness of various aspects of the program, you can sensibly decide how to do things better in the future.

Setting Criteria for Program Effectiveness (Absolute Decisions)

Criteria for Program Effectiveness (Absolute Decisions)

Before absolute decisions can be made, the evaluation data must be interpreted, and in order to do this, you must establish certain judgmental standards and values about desirable levels of performance. It is relatively easy to establish criteria to determine whether students have answered an item on a measurement device adequately. It is somewhat more difficult, or at least more a matter of judgment, to determine whether a student has achieved an objective which is measured by several items or instruments. What percentage of items must the student answer correctly in order to meet the objective? All of them? Eighty percent? Fifty percent?

It is even harder to specify what percentage of students should achieve a posttest objective in order to consider the program as having been "effective" in achieving that objective. If 75% of the students were successful, would the program be considered effective, or would 90% have to achieve that objective? Though you may want to set standards before you begin evaluating your program, it may be difficult to do this. There are no uniform guidelines, and only experience can produce confidence that the standards set are appropriate and meaningful.

Guidelines for Absolute Decisions

It *is* possible to suggest some guidelines for making absolute decisions.

1. If results indicate that a large number of one category of objectives in your program cannot be realistically and adequately met and/or if they account for frustrations and lowered motivation among students (unanticipated and undesirable side effects), a decision to delete some of the lower priority objectives is in order.

2. If, on the other hand, certain particularly desirable objectives have been achieved, continuing efforts in these areas are indicated.

3. If these desirable objectives are not being met adequately, some process revision may be in order, such as enhancing learning materials and activities.

4. If still others are met readily and completely, then perhaps a decision to redirect resources to areas where more attention is needed is in order.

5. Finally, if overall results are vague or ambiguous, further evaluation is probably needed. It is generally desirable to evaluate periodically in any event.

Weighing Results (Relative Decisions)

Weighing Results for Relative Decisions

The essence of the interpretation problem is the distinction between *educational significance* and *statistical significance*. When very large numbers of students are included in the evaluation design, very small differences in results may become statistically significant. In the real world of education or guidance, where there is competition for resources, decisions must usually be based on a more realistic estimate of significance. Depending on the number involved in the sample, the level to which the differences are found to be "significant," and the variables being measured, the evaluator and/or the decision makers must determine whether these differences are educationally significant. After an evaluation has shown that a program has produced the observed differences between experimental and control groups, someone must decide if the program did as well at helping students as was desired. In determining this, s/he is determining whether the results are educationally, as well as statistically, significant.

Statistical vs. Educational Significance

Certain generalizations regarding this process of relative decision making can be made and will perhaps make the job of decision maker and/or evaluator easier.[5]

1. When a significant difference is found favoring one program and no major differences are found in secondary variables, the more effective program will usually be adopted or expanded. For example, you tried out two different textbooks, A and B, on decision-making skills. The books cost the same and neither teachers nor students had a preference for one over the other. However, evaluation results showed that students who studied book B fared significantly better than those who worked with book A on a criterion-referenced test. Book B is subsequently adopted for all students.

2. When results reveal no statistically significant differences in the primary data (as they frequently do), decisions can be made in terms of the more secondary data. For example, if two programs are compared and found similarly effective, consideration is given to things like which approach teachers and/or students prefer, which is more economical, more convenient, etc. Back to our example of one-to-one vs. group counseling. Should a summative evaluation find one-to-one counseling and group counseling equally effective in helping students develop tentative long-range career goals, the group approach is certainly more cost-effective and probably should be adopted.

3. The secondary data can be determining factors also, where statistically significant differences are deemed to be not significant educationally. The decision to adopt a group counseling approach above might stand here too if evaluation results showed statistically but not educationally significant differences in favor of the one-to-one counseling.

4. Even when statistically and educationally significant differences are found in the primary data, a trade-off relationship often remains between the primary and secondary factors based on economics. This is where cost-effectiveness analysis pays off. How much is it worth in time, dollars, and resources to produce the likely favorable results? What is the probability of success -- short-term and long-term? Are there any undesirable side effects? An example of this kind of trade-off might be that the personal

[5]The following is based on S. Isaac and W. B. Michael, *Handbook in Research and Evaluation* (San Diego: Knapp, 1971), p. 155.

benefits of a one-to-one counseling program for all students might be offset by the cost of providing such intensive counseling services.

It is not possible to provide a recipe or cookbook approach to the process of decision making. What we have tried to do is touch on the relevant factors involved in making decisions based on evaluation results and to suggest some possible ways in which these factors may be weighed to produce a rational and practical decision in a local school setting.

Summary

Summary

In this final section of the module, we have discussed the types of decisions made as a result of summative evaluations. Criteria for program effectiveness must be set to facilitate absolute decisions, but it is a difficult task requiring judgment and experience. In making relative decisions, the data must be carefully weighed and deemed important only if they are educationally as well as statistically significant.

ACTIVITY 6 - MAKING DECISIONS BASED ON RESULTS

The purpose of this exercise is to provide you with practice in the skills needed to make decisions based on summative evaluation results.

Read the case study below and follow the accompanying instructions.

Case Study

Recall our ongoing evaluation at Lehigh. It is now a semester later, the results of the posttest are in, and all data have been processed and analyzed. Looking first at data from Adams Junior High (where the learning units were implemented), results indicated only moderate success in achievement of the eight objectives. Growth from pretest to posttest occurred on only four of them -- Objectives 2, 3, 4, and 5 -- and these were related to knowledge more than to behavior. In addition, when these data were compared with data from control students at Roosevelt Junior High, statistically significant differences (at the .05 level) occurred in favor of the Adams students on only two objectives (2 and 4, again, one from each learning unit).

In contrast to these relatively unpromising results, other, more subjective data yielded more favorable results. Teachers and counselors involved in the program at Adams reported that the incidence of hostile and agressive behaviors in the classroom was lowered. They thought this might be related to the attention being given to identifying problems and seeking help for them. There was no concrete data to back up this observation, unfortunately. Both teachers and students were surveyed to assess their reactions to the training units, and reactions from both groups were highly favorable. Teachers enjoyed teaching the units and found the learning activities highly motivating and novel, and students reported similar feelings of interest and enjoyment. Public support, on the other hand, was lukewarm; parents, on the whole, were more interested in the more academic portions of the program related to vocational information.

Administrators had no input regarding the program except that it had cost about $3,000 to implement at the school, and that an alternative, untried program covering similar objectives, but using a textbook and film format, would cost about the same.

You are Director of Guidance (still at it!) and, in conjunction with other school personnel, you need to make some decisions regarding the future of the program. Consider the results and information above, and discuss with the others of your group your recommendations (and rationale for them) for the program's immediate future. Focus your discussion on the following:

1. Should the program be continued, modified, or discontinued?

2. If it is to be modified, what changes do you recommend?

Then, working by yourself, write down your recommendations along with your rationale for them. Submit your recommendations and rationale to the Coordinator as a measure of your achievement of this outcome.

SUMMARY

Summative evaluation seeks answers to this question: Did the program meet its goals? This module takes you through six steps in conducting summative evaluations.

Step 1 is to design the evaluation. This requires specifying resources and target population and selecting outcomes to be evaluated, methods of data summary and analysis, and design type.

Step 2 is to select or develop the evaluation instrument. Appropriate measures include norm- and criterion-referenced tests and unobtrusive measures. Existing instruments may be adopted or adapted or new instruments can be created. Multiple measures should be used.

Step 3 is to administer the instrument. This involves scheduling the data collection, selecting and training instrument administrators, and orienting students.

Step 4 is to process and analyze the data. Processing can be done by hand, computer, or ADP; analyzing includes selecting the method of data presentation and calculating tests of statistical significance.

Step 5 is to relate program costs to program effects. Costs are determined by making realistic estimates, stating assumptions, and considering variables. A cost-effectiveness ratio, equal to the outcome achieved divided by the total cost of the outcome, makes it possible to compare costs and effects.

Step 6 is to make decisions based on the results of the summative evaluation. The kinds of decisions and criteria vary; the module includes guidelines for absolute and relative decisions.

The rest of the module is devoted to the Postassessment and the Application. Since you have already submitted some assessment materials to the Coordinator, the Postassessment is brief.

POSTASSESSMENT

1. List three reasons why summative evaluation is valuable and three questions it helps to answer.

2. List three major concerns involved in administering instruments.

APPLICATION

Prepare a plan and schedule for the product evaluation you will begin at the start of the next school year. The purpose is to assess students' attainment of guidance objectives. Use the attached form for developing your plan which should include the following:

1. At least three tasks required to accomplish each general phase of the evaluation (more will be needed for some phases).
2. The product that will result from each task.
3. The approximate date when you expect the task will be completed.
4. The name(s) or title(s) of the person(s) to whom responsibility for completing each task will be assigned (including outside consultants where needed).

For example:

General Phase	Task	Product	Date	Responsible Person(s)
1. Determine evaluation goals.	1. Select objectives	1. List of objectives	9-30-75	Counseling Staff & Director of Guidance
	2. Specify population	2. Statement of plan	9-30-75	Director of Guidance
	3. Specify resources	3. List of resources	9-30-75	Principals, District Superintendent

Although your planning should be independent, you may wish to consult with the Coordinator and other persons at this workshop for assistance. Begin planning at the general -- major phase -- level; then try to fill in the substeps, or tasks, within each phase. It may take more than one school to carry out your plan, depending on the scope of your effort and available resources. You may need to draft and revise your plan more than once.

When you have completed your plan, please submit it to the Coordinator.

EVALUATION PLAN AND SCHEDULE WORKSHEET

General Evaluation Phase	Tasks (Indicate at least three -- more if needed -- for each phase)	Products	Completion Date	Responsible Individual(s) (Indicate consultants if needed)
I. Determine evaluation goals.				
II. Select a design.				

EVALUATION PLAN AND SCHEDULE WORKSHEET

General Evaluation Phase	Tasks (Indicate at least thee -- more if needed -- for each phase)	Products	Completion Date	Responsible individual(s) (Indicate consultants if needed)
III. Select and/or develop instruments.				
IV. Administer instruments				

EVALUATION PLAN AND SCHEDULE WORKSHEET

General Evaluation Phase	Tasks (Indicate at least thee -- more if needed -- for each phase)	Products	Completion Date	Responsible Individual(s) (Indicate consultants if needed)
V. Process and analyze data				
VI. Determine costs and cost-effectiveness ratios.				

EVALUATION PLAN AND SCHEDULE WORKSHEET

General Evaluation Phase	Tasks (Indicate at least three -- more if needed -- for each phase)	Products	Completion Date	Responsible Individual(s) (Indicate consultants if needed)
VII. Make decisions				

75

76

APPENDIX

SELECTED LIST OF TEST PUBLISHERS

American College Testing Program, P.O. Box 168, Iowa City, Iowa 52240

American Guidance Service, Inc., Publishers' Building, Circle Pines, Minnesota 55014

Australian Council for Educational Research, Frederick St., Hawthorn E.2, Victoria, Australia

Bobbs-Merrill Company, Inc., 4300 West 62nd Street, Indianapolis, Indiana 46268

Bureau of Educational Research and Service, University of Iowa, Iowa City, Iowa 52240

California Test Bureau/McGraw-Hill, Del Monte Research Park, Monterey, California 93940

Committee on Diagnostic Reading Tests, Inc., Mountain Home, North Carolina 28758

Consulting Psychologists Press, Inc., 577 College Avenue, Palo Alto, California 94306

Cooperative Tests and Services, Educational Testing Service, Princeton, New Jersey 08540

Educational and Industrial Testing Service, P.O. Box 7234, San Diego, California 92107

Educational Test Bureau, Division of American Guidance Service, Inc., 720 Washington Avenue, S.E., Minneapolis, Minnesota 55414

Educational Testing Service, Princeton, New Jersey 98540

Guidance Centre, Ontario College of Education, University of Toronto, 1000 Yonge St., Toronto 289, Ontario, Canada

Harcourt Brace Jovanovich, Inc., 757 Third Avenue, New York, New York 10017

Houghton Mifflin Company, 110 Tremont Street, Boston, Massachusetts 02107

Institute for Personality and Ability Testing, 1602 Coronado Drive, Champaign, Illinois 61822

Lyons and Carnahan, 407 East 25th Street,Chicago, Illinois 60616

Personnel Press, Inc., 20 Nassau Street, Princeton, New Jersey 08540

The Psychological Corporation, 304 East 45th Street, New York, New York 10017

Psychometric Affiliates, Box 3167, Munster, Indiana 46321

Public Personnel Association, 1313 East 60th Street. Chicago, Illinois 60637

Scholastic Testing Service, Inc., 480 Meyer Road, Bensenville, Illinois 60106

Science Research Associates, Inc., 259 East Erie Street, Chicago, Illinois 60611

Stanford University Press, Stanford, California 94305

Stoelting Co., 424 North Homan Avenue, Chicago, Illinois 60624

Teachers College Press, Teachers College, Columbia University, New York, New York 10027

General (Relevant to all or most objectives)

Anderson, S.B., Ball, S., Murphy, R. T., and Rosenthal, E. J. *Anatomy of Evaluation.* Princeton, N.J.: Educational Testing Service, 1973.

Comprehensive encyclopedia approach to evaluation, aimed at program sponsors and evaluators. Includes articles on evaluation models, functions, and targets; program objectives and standards; typical variables; measurement approaches; reactive concerns; analysis and interpretation. A bibliography is provided for each topic.

Gottman, J. M. and Clasen, R. E. *Evaluation in Education: A Practitioner's Guide.* Itasca, Ill.: F.E. Peacock, 1972.

This book employs a workbook approach in presenting basics of educational evaluation. Seeks to develop dkills in four areas: needs assessment, developing objectives and measurement procedures, flowcharting, and designing and using quality control (evaluation) procedures.

Isaac, Stephen and Michael, William B. *Handbook in Research and Evaluation.* San Diego: Knapp, 1971.

Comprehensive, yet concise coverage of basics of research and evaluation procedures. Includes chapters on planning, designs and methods, instrumentation and measurement, statistical techniques and analysis, and criteria and guidelines for proposals and reports.

Weiss, Carol H. *Evaluation Research: Methods of Assessing Program Effectiveness.* Englewood Cliffs, N.J.: Prentice Hall, 1972.

Another comprehensive, concise look at evaluation in the area of social programs, written at undergraduate or graduate level. Practical emphasis on evaluation in the context of action programs -- e.g., schools, mental helth institutions, job training programs, etc. Covers formulating goals, unanticipated consequences, measurement variables, data collection, evaluation design, cost-benefit analysis and PPBS, problems in the action setting, and utilization of results.

Objectives 1 and 7 (Purposes of evaluation and decision making)

Baker, R. L. "Curriculum Evaluation." *Review of Educational Research,* 39, No.3, June 1969, pp. 346-358.

Portion of longer article presents review of the literature regarding the purpose and scope of curriculum evaluation, the need for objectives and instructional specifications, and definition of formative and summative evaluation. Good background source and reference to further readings.

Gilchrist, Robert S. and Roberts, Bernice S. *Curriculum Development: A Humanized Systems Approach.* Belmont, Calif.: Lear Siegler, Inc., 1974. Chapter 6: Evaluation as Communication, pp. 80-106.

Excellent discussion of the decision-making and communication needs involved in educational evaluation. Presents paradigm of curriculum development evaluation design similar to that used in these modules. Deals with process and product evaluation and integral parts of the model. Also focuses on data-gathering and measurement techniques to a more limited degree.

Objective 6 (Costs and cost-effectiveness analyses)

Kraft, Richard H. P. "Cost Effectiveness Analysis in Vocational-Technical Education." In *AERA Monograph Series on Curriculum Evaluation.* Chicago: Rand McNally, 1974, pp. 55-77. (With special emphasis on pp. 55-59 and 69-77.)

Discusses purposes, benefits, and effectiveness analysis and provides detailed case study of a vocational-technical education program where benefits could be quantified in monetary terms (i.e., future salaries). Example not as relevant to our needs where benefits are less quantifiable than future salaries.

NEA Committee on Educational Finance, *Planning for Educational Development in a Planning, Programming, Budgeting System.* Washington, D.C.: National Education Association, 1968.

Defines PPBS with special emphasis in the educational context. Offers logical and relatively simple approach to cost-effectiveness analysis by estimating costs and program outcomes, and provides an illustrative example of this approach in the context of day-care programs.

Planning, Programming, Budgeting System Manual for State of California School Districts, 2nd preliminary edition. Sacramento, Calif.: State Department of Education, 1970.

Comprehensive manual including description and implementation of operational PPBS, bibliography, and system analysis process.

TABLE OF CONTENTS

	Page
Coordinator's Role and Functions	1
Introductory Activity (optional)	3
Introductory Remarks	4
Activity Feedback and Assessment Criteria	5
Introduction to Application	15
Sample Evaluation Instruments	16

COORDINATOR'S GUIDE

Module 11

CONDUCTING SUMMATIVE EVALUATION

(Cost-Impact Studies)

Developed at the American Institutes for Research, under support by the United States Office of Education, Department of Health, Education, and Welfare under Part C of the Vocational Education Act of 1963.

August 1975

Reprinted with minor revisions, July 1978

COORDINATOR'S ROLE AND FUNCTIONS

Your role as coordinator is crucial. It may be thought of in four categories.

Set the Tone

Set the right mood. Don't make things deadly and boring. Inject humor into the activities and discussions, let people joke around and have some fun. On the other hand, make it clear that there is a very serious purpose behind it all. People should be relaxed, but alert, interested, and motivated.

Set the Pace

Maintain the right pace. If things bog down, inject some humor, ask some provocative questions, get a lively discussion going. Some sections can be summarized orally to speed things, and this can be planned ahead. If things are going too fast and people are getting lost, slow it down, let them ask questions, spend time orally covering the points. Keep the flow smooth at junctures in the module--winding up one activity with a satisfying resolution and easing participants into the next. Take breaks as you sense they are needed. Be flexible in structuring activities, adapting to individuals and situations as needed. Regard times listed in the "Module Outline" as flexible.

Facilitate

Encourage discussion and interaction from the participants. Bring out the shy people; don't let the aggressive ones dominate. Seek out questions and uneasiness, get them into the open, talk them over, especially at the beginning. Watch facial expressions and body language. Be a trouble shooter. Spot problems and work them out. In short, act as a guide through the module, but try not to get in the way.

Evaluate

Make sure participants are headed in the right direction; nudge them that way when they're not. Judge whether they perform adequately in the postassessment items, the activities which are part of the assessment (see the "Assessment Criteria") and the Application. Keep a record of how each participant does. In general, maintain the quality level of the workshop.

Specific Functions of Coordinator (to be used as a checklist as you go through the module)

Prior to workshop:

1. Study the module thoroughly. Be familiar with the participant materials and this Coordinator's Guide.
2. Make sure all needed materials are present for the workshop. This will include tape-slide equipment, a blackboard and chalk, paper and pens.

At the workshop:

1. Introduce yourself to participants, and them to each other. Briefly explain your background and the role you will play in the module.
2. Establish time limits (lunch, when day ends) and schedule for the day, and do your best to stick to them.
3. Conduct Introductory Activity (see "Introductory Activity" herein).
4. Introduce the basic purposes and structure of the module (see "Outline of Introductory Remarks"). Answer any questions.
5. Run the tape-slide presentation. Lead any discussion following this.
6. Start participants on the text. Lead discussions and practice activities as you go through the module. Provide feedback on the practice activities. (See "Assessment Criteria.") Collect evaluation data on the participants.
7. Conduct the Postassessment. Collect the results. Evaluate each of the participants on her/his performance. (See "Assessment Criteria.") Keep a written account of this.
8. Start participants on the Application. (See "Introduction to Application.") Evaluate the plan produced. (See "Assessment Criteria.")
9. Conduct a Wrap-up session. Your tasks here are to
   a. Summarize what has gone on and been accomplished.
   b. Resolve any unanswered questions.
   c. Point out sources for additional study. Go through the Reference section briefly; add any sources you know of.
   d. Mention any technical assistance available--experts related to module topics to whom participants might be able to turn.
10. Throughout, observe how things go; collect suggestions for ways to improve the module. Keep a written account of these.
11. Submit the results of 6, 7, 8, and 10 to the overall workshop director.

INTRODUCTORY ACTIVITY

(OPTIONAL)

Coordinator: Look over this Introductory Activity and determine if it is appropriate to your setting, time constraints, and participants' needs.

To introduce you to this module we want you to get together in small groups of three or four to make a decision. I will pose for you a situation, which you all have undoubtedly faced in your own day-to-day life at some time. In your groups, discuss how you would go about making the decision required and bring back your ideas to share with the rest of the group. You have about ten minutes to discuss the problem and another ten minutes to share your ideas.

Suppose that your family needed a new car and that it was up to you to decide which car to buy. How would you go about making your decision? Specifically, what questions would you ask and what data do you need before deciding? Jot down your ideas if you want, and let's get together to share each group's ideas in about ten minutes.

Feedback on Car Decision

Review the main points brought out. Ideas will probably include the following. (Share those not mentioned with participants.)

A. What are my needs? (e.g., size and type of car (sedan, wagon, van; large, small), gas and maintenance economy)

B. What are my preferences? (e.g., style, color, new/used)

C. How much can I spend?

D. What car(s) best meet my needs and preferences?

E. How much do they cost? What is the best price I can get?

F. Which is the best car for the money?

G. What other information can I get to help make the decision? (e.g., test drives, friends' recommendations, consumer reports, availability of service and parts)

H. How can I know I made the right decision after I buy a car? (Re-answer some of the above questions: Does the car meet my needs--is it roomy, economical enough? Does it drive well? Has it been free of problems? Is servicing good?)

Relationship of Opening Activity to Module

So what does all this have to do with the module? The questions you asked regarding the car are logical to ask in order to make the decision, and decisions based on the best information available are what this module is all about.

INTRODUCTORY REMARKS

Almost daily you have decisions to make for which you evaluate options and alternatives. Evaluations like this are relatively easy because most of the data are easy to get. But what if you were asked to evaluate something which many people used and the results were of vital importance to a decision to continue or discontinue use of the product? For example, your tax dollars were being used to produce an innovative school curriculum to which your kids and all the kids in your community were being exposed. You, represented by the school board, are trying to decide whether to continue the curriculum. In this case, you would probably try to be more systematic and inclusive in your evaluation than in deciding which car to buy. Because the consequences of your decision are probably greater and because more people are affected, you would aim for a more accurate and objective evaluation.

That is what this module is all about.

Outcomes and Skills (See Module Outcomes)

Have participants read through the outcomes and skills listed in their modules.

Practical Value of this Module

-- To make you aware of the need to evaluate and the possible lack of evaluation in your setting

-- To help you in making decisions

-- To help keep you accountable

-- To continue the process of developing a comprehensive career guidance program

Examples of Some Problems These Skills Can Help Solve

A. You've just implemented a new drug education program; you want to know its effects, both intended and unintended.

B. You need to develop some tests to see if students have achieved the objectives of some new guidance learning units.

C. You must decide whether or not to adopt a new series of vocational information learning units.

D. You want to find out whether your individual counseling program is effective in helping students make tentative, long-range goals.

Tape-Slide Presentation

Explain to the participants that they will now observe a tape-slide presentation which will give them an overview of Phase 4 of this planning and evaluation approach.

Outcome 1. Measurement: Postassment Item 1.

Criteria: There are of course many reasons why summative evaluation is important. The three mentioned in the module are:

1. Decision making
2. Accountability
3. Program improvement

Any other reasonable answers are acceptable also.

Summative evaluation helps to answer many questions. Those discussed in the module include

1. Tells you what objectives you have achieved
2. Measures the side effects of the program
3. Helps determine the cost-effectiveness of the program or elements of the program.

Again, any other reasonable answers are acceptable also.

Outcome 2. Measurement: Activity 1.

Criteria: 1. You do need a design for this evaluation. You need it to ensure internal and external validity. That is, you need it to be able to determine if changes in student ability are the result of the learning units and to determine whether the results you obtain here are generalizable to other situations. The latter concern is very secondary, however, as the entire district is involved in the program.

2. The budget is limited. But you do have lots of human resources to bring to bear. This is the only information provided to you.

3. The target population is all students in the district. It might also take in various adult groups--teachers, parents, and so on.

4. You probably can't measure all objectives--the program is too large. Thus you will have to choose those most important. You *will* probably want to include at least some objectives from each of the three areas: academic learning, educational/vocational, and inter/intra-personal goals. The units also focus on both knowledge and attitude outcomes, so you will probably want to include both these. Within these categories, you will have to decide what objectives are most important. One guideline: Which ones bear most heavily on decisions about maintaining or ending the program? You may also want to try to gauge side effects of the program.

5. The best design is experimental. But this presents a problem, for all students in the district will be involved. Hopefully you've used this design in the pilot tests. The only way it could be continued would be through a comparison with students in other similar districts. But pretests can be used. This will allow for comparisons and measurement of growth. While this will not be an experimental design (it will be a good descriptive design), it *will* be useful and helpful in evaluating the program.

6. A sampling strategy will probably be desirable. Including all district students in the evaluation would probably be prohibitively expensive. The best method is through use of a random numbers table as described in the text.

For each of these topics, give credit to answers which capture the essence presented here, or reasonable alternatives. For those who have failed to grasp the concepts, explain these answers and help them to

understand them. Discuss the questions as needed to make sure participants *have* grasped the concepts.

Outcome 3. Measurement: Activity 2.
Criteria: First, since items are being developed directly for the outcomes in question, criteria-referenced measurement is appropriate. The completed chart would look something like this.

Objective	Appropriate Measurement Technique	Sample Item
1. Knowledge Objective	Paper-and-pencil test is the easiest of short-answer variety, since "list" is the objective. An alternative would be an interview.	List three reasons why it is important to consider several educational and vocational alternatives in the process of setting long-range goals. a. _____ b. _____ c. _____
2. Knowledge Objective	Again, a short-answer paper and pencil instrument is easiest, perhaps of a checklist or rating-scale format. Again, an interview would be an alternative. In this case, standardized, normed instruments might also be useful. And behavioral observation (teacher, parent) might contribute useful information.	Rate each of the following from 1 (most like me) to 5 (least like me): 1. Like outdoor activities 2. Think going to church is important 3. Can do math well
3. Attitude Objective	An attitude rating scale, open-ended questionnaire, or simple checklist.	Indicate your feeling about each of the following: __ Yes __ No I would be willing to seek help from someone if I had a problem with my health.
4. Performance Objective	Unobtrusive measure. Also a questionnaire.	A simple check of the career guidance library records.
5. Attitude Objective	An open-ended questionnaire or rating scale using hypothetical situations as stimuli for the responses.	Suppose that you had just been asked by your English teacher to play the lead in the spring play. Describe in a sentence or two how you would feel.

These suggested techniques are not all-inclusive. And the example items are obviously only that--examples. Give credit to these answers and other reasonable ones, and examples that fit the techniques described.

Outcome 4. Measurement: Postassessment Item 2.
Criteria: The three concerns dicussed in the module were:

1. Scheduling
2. Selecting and training administrators
3. Orienting students

There are obviously many other concerns involved in administering the instrument. Participants may well have had a good deal of experience with this activity. Give credit for any reasonably important and sensible responses.

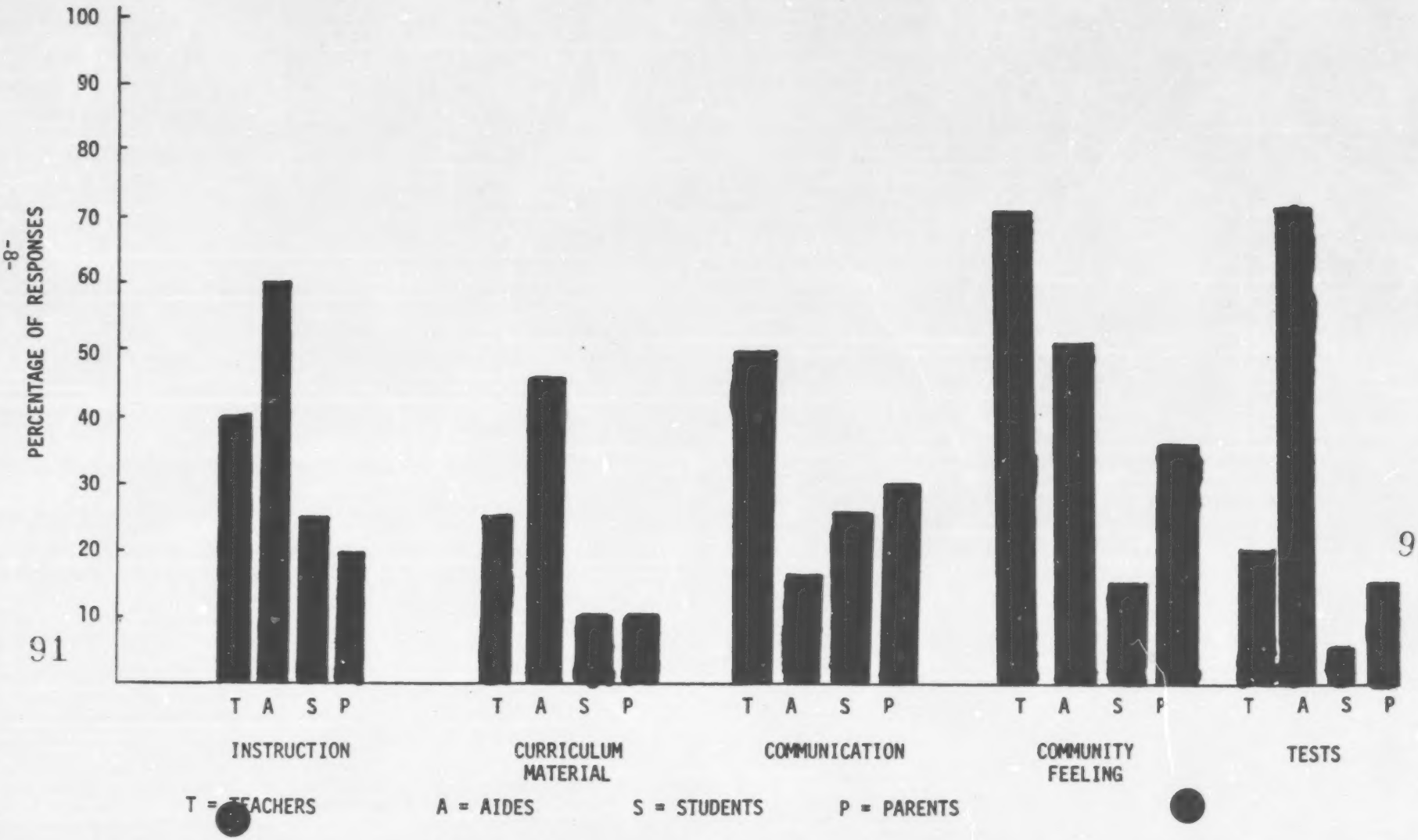
Outcome 5. Measurement: Activities 3 and 4
Criteria: Measurement of this outcome will be made as a result of participants' performance on activities 3 and 4. Activity 3 relates to presenting data while Activity 4 relates to processing and analyzing data.

For Activity 3, participants should have completed the bar graph of "excellent responses" so that the following scores for each group are presented.

Instruction:	T(teachers)	- 40
	A(aides)	- 60
	S(students)	- 25
	P(parents)	- 20
Curriculum:	T(teachers)	- 25
	A(aides)	- 45
	S(students)	- 10
	P(parents)	- 10
Communication:	T(teachers)	- 50
	A(aides)	- 15
	S(students)	- 25
	P(parents)	- 30
Community Feeling:	T(teachers)	- 70
	A(aides)	- 50
	S(students)	- 15
	P(parents)	- 35
Tests:	T(teachers)	- 20
	A(aides)	- 70
	S(students)	- 5
	P(parents)	- 15

Participants' graph should look like the graph on the following page.

BARGRAPH (HISTOGRAM) REPRESENTATION OF PERCENTAGE OF "EXCELLENT" RESPONSES BY QUESTIONS AND BY GROUPS



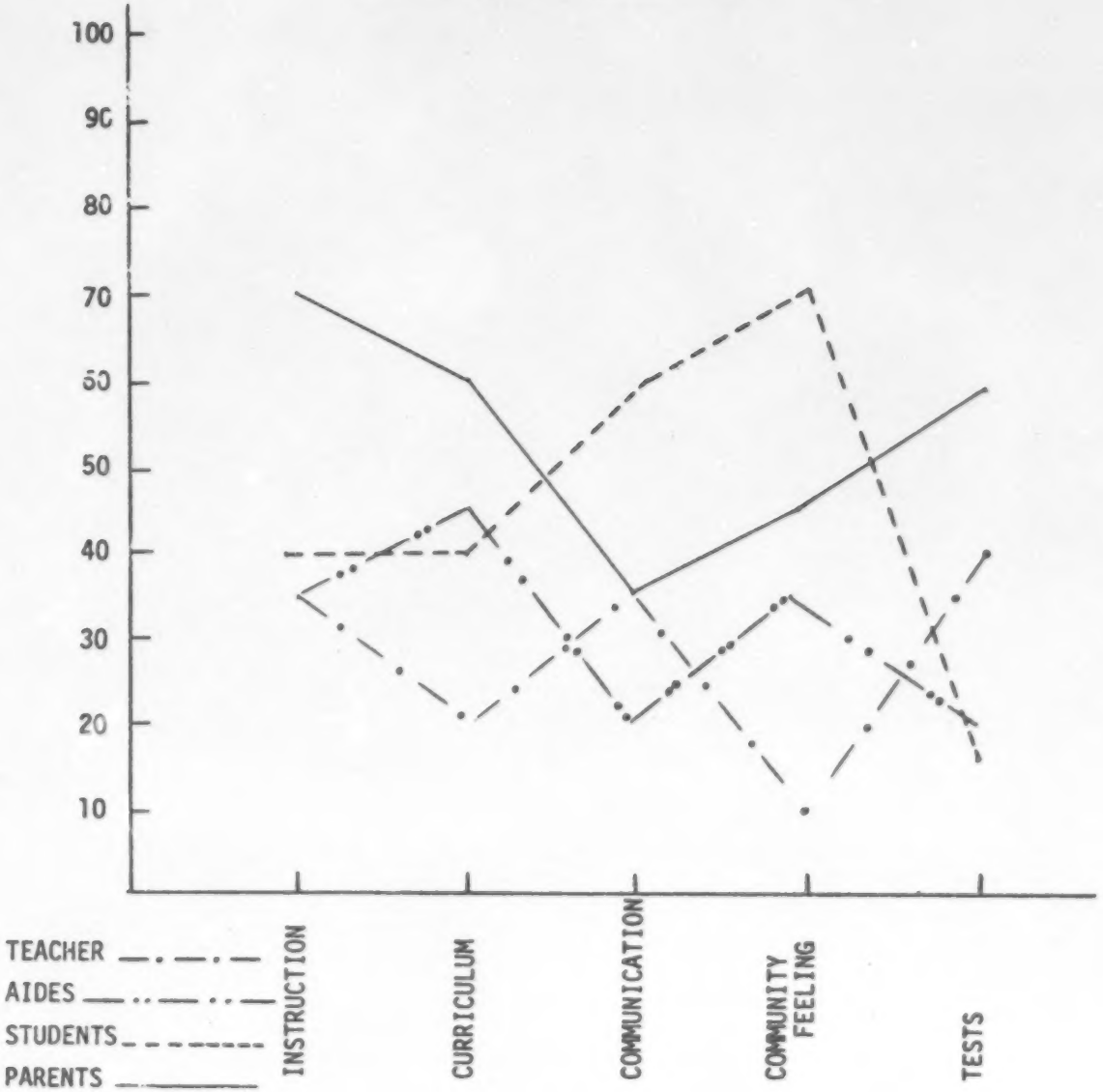
In the line graph of "average responses", the data for students are given. The scores for teachers, aides, and parents should be the following:

Teachers (—. —. —.):	instruction	- 35
	curriculum	- 20
	communication	- 35
	community feeling	- 10
	tests	- 40
Aides (—.. —..):	instruction	- 35
	curriculum	- 45
	communication	- 20
	community feeling	- 35
	tests	- 20
Parents (_____):	instruction	- 70
	curriculum	- 60
	communication	- 35
	community feeling	- 45
	tests	- 60

The participants' graphs should look like the graph on the following page.

PERCENTAGE OF "AVERAGE" RESPONSES FOR

PROGRAM COMPONENTS BY GROUPS



Activity 4. The following are suggested as the best responses to the questions on the Planning Form which participants are to complete. However, they are not the only correct answers, and responses which the participants can back up with will-reasoned answers will be acceptable.

1. Hand scoring is the most appropriate technique, considering the limited resources, available clerical help, and relatively small sample size.
2. Data can be tallied in any one (or more) of the following ways:
   (a) number of students answering each item correctly;
   (b) number of students mastering each objective. Obviously pretest, and posttest, and Adams and Roosevelt groups should be tallied separately.

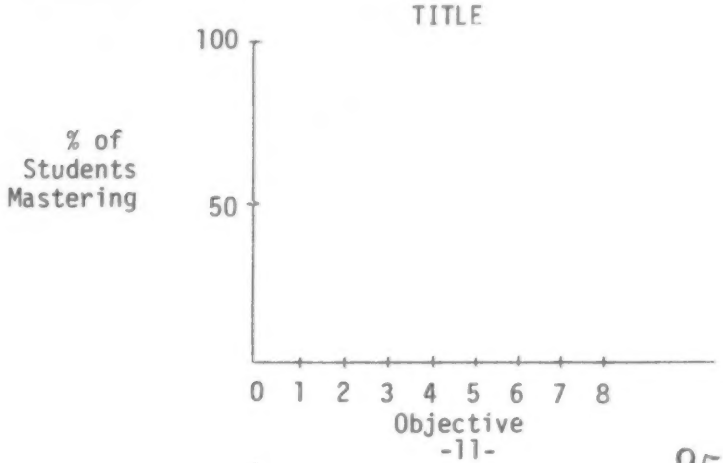
   Either (a) or (b) above or other reasonable response is adequate.
3. An indication that tables and/or graphs can be used to present data is acceptable.
   Any table or graph on the order of the following is acceptable.

   a. Table:

TITLE

Objective Number	% of Students Mastering
1 ↓ 8	

   B. Graph:



4. An answer including the two following points (and other reasonable responses) is acceptable:
   a. tests of statistical significance
   b. the statistician mentioned in the case study should be consulted, since your expertise in this area is limited.

Outcome 6. Measurement: Activity 5
Criteria: Measurement of this outcome is in terms of the effectiveness, unanticipated side effects, and cost variables which are considered for each of the three schools. An analysis of these variables in each of the schools follows.

School A

Effectiveness: Only 23% of the students turned in forms. This is low, probably because the monitoring system is a little haphazard in that it places all the responsibility for contact on the students, with no provision for follow-up by members of the career guidance team.

Unanticipated side effects: These are mixed; increased motivation was a positive side effect for some students but not for all.

Estimated costs: Low in Counselor time (3hrs. x $7.00/hr.=$21.00), but very high in materials ($25.00 + $375.00 = $400.00).

School B

Effectiveness: This is hard to rate since we do not know the size of the school. In a small school, an unstructured system such as this one would probably work well and reach all students; but in a large school, there is a high likelihood that some students would be missed by this system. The counselors report that they were able to contact 100% of the students they took responsibility for, but the problem is there may have been some students for whom no one took responsibility.

Unanticipated side effects: This system had good side effects. It increased closeness between career guidance team members and the students.

Estimated costs: Counselor time: 12 hrs. x 5 counselors x $7.00/hr.= $420.00

School C

Effectiveness: This system is good; it gets the job done, as all students do get contacted.

Unanticipated side effects: This system produced fairly serious negative side effects in that it antagonized the homeroom teachers.

Estimated costs: Counselor time: Two counselors x 1 hour x $7.00/hr.= $14.00. Also three counselors x 3 hours/week x 12 weeks x $7.00/hr. = $756.00. Total = $14.00 + 756.00 = $770.00.

Analysis

In terms of costs alone the systems would be ranked: 1. School A, 2. School B, and 3. School C. However, the benefits or effectiveness of these systems must also be considered. School A's system reached only 23% of the students. School B's system has a defect in that it might be missing some students. Correcting for this in School B may increase staff time, but it does seem that the greater effectiveness of School B's system would rank it first, followed by Schools A and C respectively. The negative side effects and high costs of C cause it to be last, even though its method for assigning students to staff members was probably the best.

Outcome 7. Measurement: Activity 6

Criteria: Assessing the adequacy of the recommendations and rationale will be a matter of judgment on the Coordinator's part. Emphasis should be on the rationale rather than the decision regarding the program's continuance per se. Among the points which should be considered are

1. The data reveal very few statistically significant differences; thus decisions should probably be made in terms of the other information available.

2. Other information favorable for the program: (a) apparent side effect of lowered aggressiveness, (b) favorable teacher reaction, (c) favorable student reaction.

3. On the other hand, only moderate public support for this part of the program causes negative pressure in terms of public accountability.

4. Because of these conflicting results, it is probably worthwhile to continue the program in somewhat modified form rather than to drop it altogether or continue it without change.

5. The program seems to have some impact in knowledge areas. Modification efforts should focus on the behavior-related efforts.

6. Modification might include an analysis of the learning materials and activities for the weaker objectives to determine where and how to beef them up. (Students and teachers might be asked for suggestions in this regard.)

7. Another untried program at similar cost is available. It should probably be tried out and compared with the current program. If results between the two programs are clearly different, the more effective program could then be adopted.

8. An overall look at the school (or district) budget is in order to see how a $3,000 expenditure (plus amount estimated to be needed for modifications suggested above) fits into the entire fiscal picture. Is achieving the eight program objectives worth the cost?

9. Any other logical consideration or point.

If respondent has considered at least five of the above points (including open-ended number 9), or very similar points, response should be judged adequate.

Outcome 8. Measurement: Application Activity product
Criteria: The time and task analysis produced should line out the tasks under each of the major headings listed, assign responsibility for each task to someone, and set a completion date for each task. In addition, the plan should be reasonably

1) Logical - do the tasks flow in logical sequence?
2) Thorough - is it detailed enough to be helpful?
3) Feasible - is it not too detailed to be burdensome; are the times allowed for the tasks reasonable?
4) Fair - are the responsibilities assigned equitably and fairly?

INTRODUCTION TO APPLICATION

Here your task is to show how this section of the module relates to the rest. To put it simply, participants should by now have learned the process one goes through in conducting summative evaluation and practiced some of the basic skills involved in this process. Now they need to figure out exactly how and when they will do this at their own setting, and who will do what.

First, review the skills and process they have learned.

1. Designing a summative evaluation
2. Selecting/developing the instruments
3. Administering the instruments
4. Processing and analyzing the data
5. Relating program costs to program efforts
6. Making decisions based on results

Now, their task is to map out a plan for doing all this in their own setting. Show participants the form to use and get them started on this activity, helping out with any questions that come up.

The criteria for the product of this effort are as follows:

The time and task analysis produced should line out the tasks under each of the major headings listed, assign responsibility for each task to someone, and set a completion date for each task. In addition, the plan should be reasonably

1) *Logical* - do the tasks flow in logical sequence?
2) *Thorough* - is it detailed enough to be helpful?
3) *Feasible* - is it not too detailed to be burdensome; are the times allowed for the tasks reasonable?
4) *Fair* - are the responsibilities assigned equitably and fairly?

On the next five pages are two sample instruments, the Module Performance Record and the Evaluation Questionnaire for Staff Development Workshops. You may wish to use these instruments to gather information for evaluating any workshop in which you administer this module, and for making decisions about future workshops. The Module Performance Record (MPR) is a form for tallying participants' achievement of objectives. The Evaluation Questionnaire seeks participants' opinions on four dimensions: (1) perceived value of the workshop; (2) effects of participating in the workshop; (3) role and performance of the coordinator; and (4) recommended improvements in the workshop. As it now stands, the questionnaire should take participants 10-20 minutes to complete. You, as module coordinator, should complete the MPR form based upon the results of the postassessment or other evidence supplied by participants. If you duplicate the Evaluation Questionnaire for participants to complete, we suggest you print it as a four page booklet.

NATIONAL CONSORTIUM ON COMPETENCY-BASED STAFF DEVELOPMENT

MODULE PERFORMANCE RECORD

MODULE TITLE: _____

WORKSHOP DATES: _____

WORKSHOP COORDINATOR(S): _____

_____

Participants' Names (Alphabetically)	OBJECTIVES (Place a check (✔) mark for each objective achieved.)						
	1	2	3	4	5	6	7
1.							
2.							
3.							
4.							
5.							
6.							
7.							
8.							
9.							
10.							
11.							
12.							
13.							
14.							
15.							

Developed at the American Institutes for Research, under support by the United States Office of Education, Department of Health, Education, and Welfare.

NATIONAL CONSORTIUM ON COMPETENCY-BASED STAFF DEVELOPMENT

EVALUATION QUESTIONNAIRE FOR STAFF DEVELOPMENT WORKSHOPS

Your responses to the brief questions in this booklet will help us evaluate the workshop you just completed and make decisions regarding future workshops. Please take 10-20 minutes to answer honestly and thoughtfully. You need not sign your name, but we do need your help. Please answer each question. Thank you.

Name (Optional) _____ Date _____

Module Title _____

A. General Issues Related to the Workshop

Respond by checking the column (A, B, C, D, or E) of the statement which best expresses your feeling or opinion on each item in the following list. If none of the possible choices precisely represents your view, pick the one that comes closest.

STATEMENTS	Column A I strongly agree	Column B I agree	Column C I disagree	Column D I strongly disagree	Column E I don't know. I have no basis for answering.
1. This workshop made a valuable contribution to my professional development.					
2. I acquired no new knowledge during this workshop.					
3. I would not recommend this workshop to anyone else.					
4. I am glad I attended this workshop.					
5. In this workshop, I experienced at least one positive change in my knowledge, attitudes, or skills.					
6. As a result of this workshop, I expect that I will help improve the career guidance program in my work setting.					
7. I experienced at least one negative effect from this workshop.					
8. This module must be improved.					
9. The Coordinator was an asset to this workshop.					
10. The Coordinator was unprepared.					
11. The Coordinator was poorly organized.					
12. The Coordinator was clear and to the point.					
13. The material and activities in this workshop were not helpful.					
14. The material and activities were applicable to my needs.					
15. The material and activities were routine and boring.					
16. The workshop's objectives addressed the training needs that I hoped they would.					

Developed by the National Consortium on Competency-Based Staff Development in cooperation with the American Institutes for Research, under support by the United States Office of Education, Department of Health, Education, and Welfare under Part C of the Vocational Education Act of 1963.

B. Please list and briefly describe up to three major positive changes that you have experienced in your knowledge, attitudes, or skills because of this workshop. Continue on the back of this booklet if necessary. If you did not experience any positive changes, please check the appropriate space.

_____ There were no positive changes.

C. Please list and briefly describe any negative effects you have experienced because of this workshop. Continue on the back of this booklet if necessary. If you did not experience any negative effects, please check the appropriate space.

_____ There were no negative effects.

D. Please list and briefly describe any improvements you anticipate in your career guidance program as a result of this workshop. Continue on the back of this booklet if necessary. If you don't expect any improvements as a result of this workshop, please check the appropriate space.

_____ I don't expect any improvements in my career guidance program as a result of this workshop.

E. Please list and briefly describe any other comments on this workshop, criticisms of it, or suggestions you have for improving it. We are especially interested in your ideas on topics or activities that should receive more or less emphasis. Continue on the back of this page if necessary.